Introduction to Econometrics

Vipul Bhatt

2023-08-21

Contents

Preface

These lecture notes are prepared for an introductory course in Econometrics. These notes borrow heavily from the teaching material that I have developed over several years of instruction of this course at the James Madison University.

One of my main objectives is to develop a textbook in Econom terics that is more accessible to undergraduate students than standard textbooks available in the market. Most of these textbooks in my opinion are densely written and assume advanced mathematical skills on the part of our students. Further, I have also struggled with their topic selection and organization. Often I end up not following the chapters in order and modify content (by adding or subtracting) to meet my students needs. Such changes causes confusion for some students and more importantly discourages optimal use of the textbook. Hence, this is an undertaking to develop a primer on Econometrics that is accessible, follows a more logical sequencing of topics, and covers content that is most useful for undergraduate students in business and economics.

Note: These notes have been prepared by me using various sources, published and unpublished. All errors that remain are mine.

Introduction

Econometrics as a sub-field of Economics is primarily concerned with statistical analysis of economic data. Consider an important economic relationship, namely, the effect of education on labor market wage. Economic theory suggests higher education should lead to greater wages. Econometrics provides a toolkit for empirically testing this relationship using real world labor market data. Note that there are two distinct analyses one maybe interested in this context:

1. What is the causal effect of education on wages?
2. Given data on education, what can we say about wages?

Before answering these questions, we need to consider several ancillary issues that are especially unique to economic relationships. For example, what kind of data can be used to answer above-mentioned questions? What are some of the specification issues we must account for in estimating the effect of education on wages? The key issue is that although economic theory provides a qualitative basis for an economic relationship, it is often a poor quantitative guide. For instance, economic theory will tell us that higher education leads to greater wages but does not tell us by how much or whether such a relation is linear or non-linear in nature.

An important distinction which is often not emphasized in most Econometrics textbooks is between two main purposes of economic analysis:

1. Explanatory analysis: here the goal is to obtain a causal effect of an **independent or explanatory** variable on a **dependent or outcome** variable of interest. In the context of our economic example, we can ask the following question: what is the effect of education on wages? Implicit here is an assumption that education has a causal effect on wages of an individual. This is inherently a quantitative question and we need a numerical answer to this question. Further, because in practice we use sample data for answering such questions, use of different samples will provide different estimates of the effect of education on wages. As a

result we need to also provide some measure of how precise our numerical estimate of the causal effect is.

2. Predictive analysis: here the goal is to obtain an accurate forecast or prediction of future (new) values of the dependent variable, given data on independent variable(s). For instance, if we are given data on education what can we say about expected wages? An answer to this question does not require a causal relationship between education and wages. The goal of predictive analysis is to forecast the dependent variable for new observations given data on the independent variable(s). These new observations could be across sample units (e.g., individuals, firms, countries) or over time (e.g., week, month, year). A more common use of predictive analysis is to generate forecasts of the variable of interest over time. Such analysis can be done by only utilizing historical data on the dependent variable and/or by using historical data on other relevant variables. What is important is that a lack of causal relationship between variables does not necessarily reduce the forecast accuracy of a predictive model.

Although the *classical regression model* is used for both types of analyses, it is important to distinguish between, for example, an attempt to tease out the causal effect of education on wages versus predicting wages based on education. Let us use two potential issues that one may face in this context to illustrate their differential treatment depending on whether our objective is to obtain a causal effect of education on wages or predict wages using data on education.

a. It is reasonable to argue that an individual's innate ability can simultaneously affect both her education and wages. However, because it is very difficult to measure innate ability, we exclude it from our analysis.

b. It is possible that dispersion in wages decreases in education so that education provides a more accurate prediction of wages for observations with higher level educations when compared to those with lower level of education.

If our goal is to obtain a causal effect of education on wages, excluding innate ability will confound the effect of education on wages. The idea here is that individuals with higher innate ability are also more likely to obtain higher education. Hence, if we exclude ability from our model, we cannot tease out the causal effect of education on wages because only part of the estimated effect from such a model can be attributed to education. This is the *omitted variable* problem that constitutes one of the most important obstacles in causal analysis is the possibility of omitting relevant explanatory variables from. Going back to our example, any causal analysis of education on wages needs to devote substantial attention to such confounding effects of omitted variables. On the other hand, the issue of wage dispersion being systematically affected by education

can only affect the precision of the estimator. Hence, it is commonplace to account for such a pattern by simply scaling the standard errors of the estimated coefficient of education appropriately.

In contrast, if our goal is to predict wages based on education, lack of data on innate ability and the consequent non-casual interpretation of estimator is not important. As long as education provides useful information about wages in data it will help improve the forecast of wages. However, a greater emphasis in this case needs to placed on the second issue of wage dispersion decreasing in education. Such a pattern in data is valuable for improving forecast accuracy and must be explicitly modeled along with the effect of education on wages for obtaining a more accurate forecast.

The aforementioned distinction between predictive and explanatory analyses is a recurring theme in this textbook. For every topic covered in the text, I will attempt to establish a link between the main issues and how their relevance/importance will vary depending on whether our goal is to predict an economic variable or explain the causal factors that determine this variable.

Chapter 1

Simple Regression Model

In this chapter we consider the simple case of establishing an empirical relationship between two variables. For simplicity, we will denote the outcome variable (or dependent variable) by Y and the explanatory variable (or independent variable) by X. Throughout this chapter we will ignore the limitations of this model and assume that all the assumptions needed for this kind of model are met in our data. In subsequent chapters we will learn how this model can be extended to incorporate real world issues we face when conducting quantitative economic analysis.

1.1 Statistical foundation of the simple regression model

A simple economic relationship often requires us to specify the population of interest and variables from this population that we wish to empirically examine. For example, we may be interested in finding out how college major choice affects starting salary of graduates. Here, the population of interest is college graduates and the variables representing this population are starting salary and college major. In this example, the outcome variable is starting salary (denoted by Y) and the explanatory variable is college major choice (X). A simple regression model addresses the following question: how does changes in X affect Y? In our example, how does differences in college majors between individuals relate to differences in their starting salaries. This is the conditional expected value of Y given X, denoted by $E(Y_i|X_i)$ and is known as the **population regression function (PRF)**. Because we usually work with a sample, the population regression function is unknown to us and our goal is to estimate this conditional mean using a sample data on Y and X.

Before attempting to estimate the PRF, we have to first make an assumption about the functional form for this function. For example, we can assume that:

$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i$$

In this case we are assuming that changes in X affects expected value of Y in a linear fashion (constant slope, given by β_1). Note that we can accommodate any kind of functional relationship between Y and X by simply changing the functional form. For example, a quadratic specification is given by:

$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i^2$$

Here, we assume that changes in X affect Y in a non-linear fashion. In this sense, the simple linear regression model is flexible enough to capture the relationship between X and Y. However, this also means that getting the functional form wrong will introduce a **specification error** in our estimation. Depending on whether we want **explain** Y using data on X or **predict** Y using data on X, the consequences of the specification error could be less or more severe. Here it is also important to emphasize the difference between **linear in parameters** vs **linear in variables**. Both of the assumed functional forms are linear in parameters β_0 and β_1. However, the first specification is linear in X whereas the second specification is non-linear in X. In this course, we will only focus on specifications that are linear in parameters.

Once we have assumed a functional form for the PRF, the next step is to estimate the model parameters of interest: intercept (β_0) and slope (β_1). These are unknown parameters that characterize the relationship between Y and X. If we knew their values, then using data on X we can compute the model value of Y. For example, suppose $\beta_0 = 5$ and β_1=2. Then, if X=10, we get:

$$E(Y_i|X_i = 10) = 5 + 2 \times 10 = 25$$

Note that the above value of Y is what our model expects Y to be when X takes a value of 10. In data, we may find that the actual value of Y corresponding to X=10 is different. This difference between data on Y and model-generated value of Y is called the **regression error term** and is denoted by ϵ_i. Formally,

$$\epsilon_i = Y_i - E(Y_i|X_i)$$

or equivalently,

$$Y_i = E(Y_i|X_i) + \epsilon_i$$

By substituting our assumed functional form for $E(Y_i|X_i)$ in the above equation we obtain the simple regression model. For example, if we assume that $E(Y_i|X_i) = \beta_0 + \beta_1 X_i$, then the corresponding regression model is given by:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{1.1}$$

The specification in is the simple regression model with parameter β_1 capturing the effect of changes in X on Y:

$$\beta_1 = \frac{dY_i}{dX_i}$$

So if X changes by 1 unit, then Y changes by β_1 units. Under certain assumptions that we will discuss later in the course, β_1 is also the causal effect of X on Y.

1.2 Ordinary Least Squares (OLS)

The goal of any empirical analysis is to find a **functional** relationship between the outcome (or dependent variable) and the explanatory (or control variable). As discussed in the previous section, the PRF is one such function. Once we have estimated this function it can be used for either explaining variations in outcome caused by variations in control or to predict the outcome given data on control. For example, economic theory tells us that consumption and income are positively related implying there must be a numerical mapping between these two variables. For example, assuming a linear relationship between consumption (Y) and income (X) we get the following PRF:

$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i$$

As discussed earlier, this relationship will not be perfect in data. This could be due to other factors affecting consumption such as gifts, inheritances, wealth etc. The regression error term captures all such influences on consumption that stem from non-income sources. This error can be negative or positive and is captured in the following regression model:

$$Y_i = E(Y_i|X_i) + \epsilon_i$$

Or equivalently,

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

If we knew the error term for each observation, then given data on consumption and income, we can simply choose values of the intercept and slope that best fits our data. However, by definition the regression error is a random variable whose values we do not observe. Consequently, the two parameters of the model, β_0 and β_1 are also unknown to us. To circumvent this problem, we focus on the sample counterpart of the regression error term called the **residuals** which is denoted by e_i and can be computed using data on the outcome and the control variable. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the sample estimators of β_0 and β_1. Let $\hat{Y}_i$ denote the **predicted** value of consumption given by:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

The predicted value of consumption is also the sample estimator of $E(Y_i|X_i)$. Now, the residuals are defined as the difference between actual consumption and consumption predicted from our model:

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

Given data on consumption and income, we can now choose values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that will minimize our residuals in some sense. This is displayed below in Figure 1.1 below:

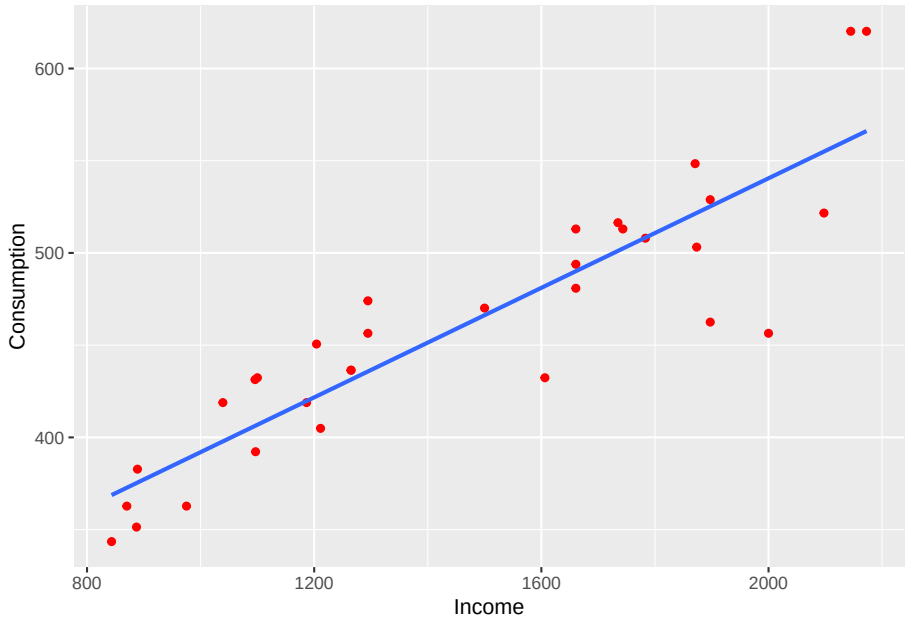


Figure 1.1: Line of best-fit

In Figure 1.1, each red dot denote the level of income and corresponding level of consumption for that particular observation. The blue solid line is our estimated regression line or $\widehat{Y}_i$. The gap between the red dot and the regression line is the residual for that observation. In this graph we hope that the regression line passes through our data as closely as possible. In other words, we would like our residuals to be as small as possible.

The canonical estimation method used in Econometrics for finding the line of best fit is called the **Ordinary Least Squares (OLS)**. Here, the values of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are chosen such that the sum of squared residuals is minimized. Let RSS denotes sum of squared residuals which is given by:

$$RSS = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2$$

Mathematically, OLS solves the following optimization problem:

$$min_{\widehat{\beta}_0, \widehat{\beta}_1} \quad \sum_{i=1}^{N} RSS$$

The solution to the above optimization problem gives us the following formulas:

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$$

and

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{N} (Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^{N} (X_i - \overline{X})^2}$$

where $\overline{Y}$ and $\overline{X}$ denote sample means of consumption (Y) and income (X). By definition, these OLS estimators satisfy the following desirable properties:

1. Sum of residuals is zero, i.e., $\sum_{i=1}^{N} e_i = 0$.
2. Residuals and the explanatory variables are uncorrelated, i.e, $\sum_{i=1}^{N} e_i X_i = 0$.

Example 1.1. Suppose you have the following data:

OLS Estimation Example

ID	Y_i	X_i
1	10	3
2	8	6
3	12	3
4	14	8

Suppose we assume that the relationship between Y and X can be captured by the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

We can now use the data given to us and compute the OLS estimators for two regression parameters. First, note that the sample means are given by: $\overline{Y} = 11$ and $\overline{X} = 5$. Second, we compute the slope cofficient by utilizing the information provided in the table below:

OLS Estimation Example contd..

ID	Y_i	X_i	$Y_i - \overline{Y}$	$X_i - \overline{X}$	$(Y_i - \overline{Y})(X_i - \overline{X})$	$(X_i - \overline{X})^2$
1	10	3	-1	-2	2	4
2	8	6	-3	1	-3	1
3	12	3	1	-2	-2	4
4	14	8	3	3	9	9
Total			0	0	6	18

Using the OLS formular, we get:

$$\hat{\beta}_1 = \frac{6}{18} = 0.33$$

$$\hat{\beta}_0 = 11 - .33 \times 5 = 9.33$$

Hence, our model provides us the following equation for the predicted value of Y_i:

$$\hat{Y}_i = 9.33 + 0.33 \times X_i$$

Using the above equation we can easily find what our model predicts about the dependent variable given data on X. Note that we can compute this predicted

value for observations for which we have data on X and compare it with the actual data on Y. The difference between the two gives us the regression residuals. This is provided in the table below:

OLS Estimation Example contd..

ID	Y_i	X_i	$\hat{Y}_i$	e_i
1	10	3	10.32	-0.32
2	8	6	11.31	-3.31
3	12	3	10.32	1.68
4	14	8	11.97	2.03

Using data used in estimating the regression model to compute predicted values of the outcome variable is called **in-sample** prediction. In the above table, we have in-sample prediction for the outcome variable. We can also similarly compute the predicted value for any value of X even outside our given sample. Such an exercise is called **out-sample** prediction. So for example, consider $X = 7$ which is not part of our sample. We can compute the corresponding predicted value of Y is given by:

$$\hat{Y} = 9.33 + 0.33 \times 7 = 11.64$$

This is one of the primary uses of a regression model, namely, the ability to predict the value of the outcome variable, given information on the value of the control variable. Another objective of estimating a regression model is to explain how X affects Y. In our example, a 1 unit change in X will change Y by 0.33 units. Under some assumptions, this denotes the causal effect of X and Y. The sign and the size of this effect provides meaningful information on the nature of the relationship between the outcome and the control variables.

1.3 Goodness of fit

After estimating a regression model using OLS a natural question to ask is how good is our estimated model in fitting data on the outcome variable. One measure of this is the **goodness of fit** which is based on the proportion of total variation in the dependent variable that can be explained by our model. A better fit would imply greater amount of variation being explained.

Essentially, our goal is to decompose the total variation in the dependent variable into two components, nameley, **explained variation** and **residual variation**. Formally,
rewriting the definition of residuals to get the following relationship:

$$Y_i = \hat{Y}_i + e_i$$

Subtract mean of the dependent variable from both sides we get:

$$Y_i - \overline{Y} = \hat{Y}_i - \overline{Y} + e_i$$

Finally, squaring and summing across all observations gives us the following identity:

$$\underbrace{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}_{TSS} = \underbrace{\sum_{i=1}^{N}(\hat{Y}_i - \overline{Y})^2}_{ESS} + \underbrace{\sum_{i=1}^{N} e_i^2}_{RSS}$$

where TSS stands for total sum of squares and measures variation in the dependent variable, Y_i. ESS denotes explained sum of squares and measures variation in the predicted value of the dependent variable, $\hat{Y}_i$. Finally, RSS, as previously defined captures variation in regression residuals. We can now define a measure of fit known as the **R-squared** which is denoted by R^2 and is defined as the ratio of ESS to TSS. Formally,

$$R^2 = \frac{ESS}{TSS}$$

Hence, R^2 tells us the fraction of total variation in the dependent variable that can be explained by our model. If $R^2 = 0$, then our model explains none of the variation in Y_i. Similarly, $R^2 = 1$, then our model explains all of the variation in Y_i. As a result, R^2 will always be a number between 0 and 1 with a higher number indicating a better fit. Note that by definition:

$$1 - R^2 = \frac{RSS}{TSS}$$

As a result, $1 - R^2$ gives us the unexplained variation. For example, a value of 0.4 will imply 40% of the variation in Y_i is explained by our model or 60% of the variation is not explained.

1.4 Applications of the simple regression model

1.4.1 Aggregate consumption function

Suppose you are interested in finding out the relationship between consumption and income at the national level. The first step in any empirical anaylsis is to

determine the empirical measures of the variables of interest, consumption and disposable income here. We will use real personal consumption expenditure as a measure of C and real personal disposable income as a measure of Y^D. Both of these data are available at the monthly frequency from the FRED Stat database and can be downloaded from the following links:

i) https://fred.stlouisfed.org/series/PCEC96

ii) https://fred.stlouisfed.org/series/DSPIC96

We will use the data from Jan-2002 through June 2019. The next step is to formulate our regression model. Here, we can use economic theory and propose a linear relationship between consumption and income:

$$E(Consumption_t|Income_t) = \beta_0 + \beta_1 Income_t$$

In macroeconomics, the above relationship is known as the **Keynesian** consumption function. John Maynard Keynes proposed that at the aggregate level, consumption changes in proportion to changes in disposable income:

$$C = a + b\ Y^D$$

Here C denotes private consumption expenditure and Y^D denotes post-tax or disposable income. a is the intercept and captures the part of consumption that is indpenedent of income. b is the slope and measures the unit change in consumption caused by a unit change in disposable income. b measures the marginal propensity to consume and is a parameter of interest we would like to estimate using data.

The implied regression model is given by:

$$Consumption_i = \beta_0 + \beta_1 Income_i + \epsilon_i$$

Our parameter of interest is β_1 which maps to the marginal propensity of consume in the original regression model. Table 1.4.1 below presents a summary of the OLS estimation results for the above model. A coefficient of 0.82 for income implies that for every dollar increase in disposable income, consumption increases by 82 cents. Under certain assumptions which will be discussed in Chapter 5, this is the **causal** effect of change in income on consumption at the aggregate level. Note that this relationship is estimated using time series data and does not tell us anything about how changes in income across individuals

affect their consumption spending. We also note that the R-squared is 0.986 which means that roughly 98 percent of variation in the aggregate consumption expenditure from one month to another can be explained by personal disposable income.

OLS Estimation of Keynesian Consumption Function

Table 1.4: OLS Estimation of Keynesian Consumption Function

	Dependent variable:
	Real Personal Consumption Expenditure
Intercept	1,205.284***
	(82.638)
Real Personal Disposable Income	0.805***
	(0.007)
Observations	210
R^2	0.985
Adjusted R^2	0.985
Residual Std. Error	128.468 (df = 208)
F Statistic	14,099.740*** (df = 1; 208)
Note:	*p<0.1; **p<0.05; ***p<0.01

Note that we can use our estimated regression model to predict future aggregate consumption expenditure. Our regression used data from Jan 2002 through June 2019. Suppose, we believe that in Jan 2020, the real personable disposable income will be 1% higher than the June 2019 level of $15007.5 Billion. Then, using our estimated model, the real personal consumption expenditure will be:

$$Consumption_{Jan,2020} = 1037.389 + 0.82 * (15007.5 \times 1.01) = \$13475.29 \ Billion$$

Note that instead of assuming a single value for the future income, we can consider a range of scenarios and compute the predicted consumption expenditure for each scenario. These scenarios are based on our understanding of the state of the economy and policy environment. For example, what if there is a tax cut that increases disposable income? Or what if there is a crash in the housing market leading to a large drop in disposable income for many households? This type of exercise is known as **scenario-based** forecasting. Table 1.4.1 below provides future consumption expenditure for these 3 scenarios.

OLS Estimation of Keynesian Consumption Function

Scenario of Jan 2020	Real Personal Disposable Income	Real Consumption Expenditure
Average income growth (Apr-Jun, 2019)	0.19%	$13375.61 Billions
One std. dev. below average	0.07%	$13360.84 Billions
One std. dev. above average	0.31%	$13390.38 Billions

1.4.2 Returns to education

An important research question in labor economics is how education affects labor income. Economic theory suggests that education makes a worker more productive through acquisition of skills that are rewarded in the labor market as higher wages. To test this in data, we would need to collect information on education and wages for a sample of workers. Often such an exercise is conducted using a cross-sectional data at a point in time. One such source is the Annual Social and Economic Supplement (ASEC) of the Current population survey (CPS) which is a monthly survey of households in the U.S. conducted by the Bureau of Labor Statistics (BLS). The data is publicly available in a user-friendly format at:

https://cps.ipums.org/cps/

We will use a sample of 1000 observations from the ASEC 2018 in this exercise. Suppose, our theoretical model for the average relationship between income and education is log-linear:

$$E[ln(wage_i)|education_i] = \beta_0 + \beta_1 education_i$$

The above formulation is a simplified version of the specification that has been extensively used in the labor economics literature. The above log-linear relationship between wages and education was made famous by Jacob Mincer and is commonly known as the **Mincerian earnings function**. The orginal specification also includes controls for years of experience. Using Census data from 1950 and 1960, he caluculated that every additional year of education increases annual earnings by 5 to 10%.

The implied regression model is given by:

$$ln(Wage_i) = \beta_0 + \beta_1 Education_i + \epsilon_i$$

where $Wage_i$ is the annual wages and salaries of individual i in dollars. $Education_i$ denotes years of education of individual i. Table 1.4.2 below presents the estimation results. A coefficient of 0.106 for education implies that for every additional year of schooling, annual wages increase by 10.6%. Again, only under certain assumption this captures causal effect of education on wages. We also note that the R-squared is 0.122 which means that roughly 12 percent of the variation in wages across individuals can be explained by differences in their education levels.

OLS Estimation of Earnings Function

Table 1.6: OLS Estimation of Earnings Function

	Dependent variable:
	Log of Annual Wages
Intercept	9.259***
	(0.135)
Years of Schooling	0.106***
	(0.009)
Observations	1,000
R^2	0.122
Adjusted R^2	0.121
Residual Std. Error	0.748 (df = 998)
F Statistic	138.887*** (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Notice how the interepretion of the coefficient of education is percent terms here: multiply the slope coefficient by 100 and interpret as implying a 10.6% increase in wages. This is because the dependent variable is transformed into natural logs. Mathematically, 100 times one unit change in natural logs of a variable is approximately equal to a percent change in the level of the variable. More on this in the next section.

1.5 Use of natural logs and interpretation of the slope coefficients

Suppose we have data on two variables: Y and X. Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

There are two things to remember when use a model like above:

1. We assume that there is a linear relationship between Y and X. This is what we mean by **functional form** or **specification** of our regression model.

2. The interpretation of the slope coefficient is in terms of original units of our data. So, if X denotes number of hours studied and Y denotes test score in points from 1 through 100, then an additional **hour** of study will change your test score by β_1 **points**.

Once we transform our data by using some form of mathematical operation we need to change both our belief about the functional relationship between Y and X, as well as our interpretation of the slope coefficient. One of the most commonly used transformation in economic analysis is the use of natural logs. For example, consider the following specification:

$$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + \epsilon_i$$

In this case, we believe that after transforming our data to natural logs, the relationship becomes linear. However, the above regression model implies that the relationship between Y and X in original units is non-linear. Specifically to obtain the above regression model, the underlying relationship between Y and X in original units must be:

$$e^{Y_i} = e^{\beta_0} X_i^{\beta_1} e^{\epsilon_i}$$

Further, the interpretation of β_1 is now different as well. Specifically, if ln(X) changes by 1 log point then ln(Y) changes by β_1 log points. A more intuitive interpretation is that if X changes by 1% then Y will change by β_1%.

To understand this section, we need to revisit the concept of natural logs. The natural logs of a number x is its log using the mathematicl constant $e \approx 2.718$ as the base. The natural logs of x is the power to which e has to be raised to be equal to x For example, natural log of 10 is 2.302585 because $e^{2.302585} = 10$. Some useful properties of the logs that are applicable to natural logs as well are:

1. $ln(1) = 0$
2. $ln(x \times y) = ln(x) + ln(y)$
3. $ln\left(\frac{x}{y}\right) = ln(x) - ln(y)$
4. $ln(x^a) = a \times ln(x)$
5. $ln(e^x) = e^{ln(x)} = x$
6. If $y = ln(x)$, then $\dfrac{dy}{dx} = \dfrac{1}{x}$
7. $\Delta ln(X) \times 100 \approx$ percent change in X. This implies that $\dfrac{dln(Y)}{dln(X)} = \dfrac{\text{percent change in Y}}{\text{percent change in X}}$

There are three types of log transformations we can use for a simple regression model. The table below shows all three with corresponding interpretations for the slope coefficient.

Model Name	Model Specification	Interpretation of β_1
Log-Log	$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + \epsilon_i$	If X changes by 1%, then Y changes by β_1%
Log-Linear	$ln(Y_i) = \beta_0 + \beta_1 X_i + \epsilon_i$	If X changes by 1 unit, then Y changes by $100 \times \beta_1$%
Linear-Log	$Y_i = \beta_0 + \beta_1 ln(X_i) + \epsilon_i$	If X changes by 1%, then Y changes by $\frac{\beta_1}{100}$ units

If we want to compare goodness of fit of two different models, then the units of the dependent variable must be the same. In the above example, we can compare the R^2 of the log-log and log-linear models as they both have natural log of Y as the dependent variable. However, neither can be compared with linear-log model. This is because a linear-log model explains observed variation in the orginal units of Y. In contrast the other two models explain variations in natural logs of Y.

1.6 Hypothesis testing and confidence interval estimation

We can use the sampling distribution of the OLS estimator to carry out hypotheses about the true population parameter for interest. For example, suppose we have the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Using the OLS we can obtain $\hat{\beta}_0$ and $\hat{\beta}_1$, both of which are computed for a given random sample. As we change our sample, we get a different set of values for $\hat{\beta}_0$ and $\hat{\beta}_1$. Hence, we should think of these OLS estimators as random variables. In principle, we can collect large number of samples and compute $\hat{\beta}_0$ and $\hat{\beta}_1$ for each sample, giving us a sampling distribution of these estimators. Let us focus on the slope parameter's OLS estimator, namely, $\hat{\beta}_1$. Using repeated samples for Y and X, we can obtain a distribution of values for $\hat{\beta}_1$ with a mean of $E(\hat{\beta}_1)$ and a variance of $Var(\hat{\beta}_1)$. Under some conditions, we can show that $E(\hat{\beta}_1) = \beta_1$ and

$$Var(\hat{\beta}_1) = \frac{Var(e_i)}{\sum_{i=1}^{N}(X_i - \overline{X})^2}$$

where $Var(e_i) = \frac{\sum_{i=1}^{N} e_i^2}{N-2}$ denotes the variance of the regression residuals.

In order to conduct hypotheses about the slope parameter, we need to make some assumption about the regression error term, ϵ_i. We will assume that the error term is normally distributed with a mean of 0 and variance of σ^2. Then, we can test any hypothesis about β_1 using the t-test. The formal procedure for a two-sided test is outlined below:

1. Specify the null and the alternative hypotheses:

$$H_0 : \beta_1 = R$$
$$H_0 : \beta_1 \neq R$$

where R can be any numerical value.

2. Compute the t-ratio:

$$t = \frac{\hat{\beta}_1 - R}{s.e.(\hat{\beta}_1)}$$

where $s.e.(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)}$

3. Under the null hypothesis, and assuming that the error term is normally distributed, we can show that the above t-ratio follows a t-distribution with $N-2$ degrees of freedom. The decision rule is given by:

$$|t| > t_c, \ reject \ H_0$$

where t_c is the critical value that can be obtained from the t-distribution table for a given level of significance.

Test of Statistical Significance: If we use $R = 0$, then the test becomes:

$$H_0 : \beta_1 = 0$$
$$H_0 : \beta_1 \neq 0$$

If the null hypothesis is true, then there is no statistically significant effect of X on Y. Most statistical softwares provide the t-ratio and the associated p-values for this test by default. Note that the results of this test cannot be used for inferring the economic significane of the effect of X on Y. For that the magnitude of $\hat{\beta}_1$ is more informative.

Example 1.2 (Test of Statistical Significance)**.** Let us go back to our example of returns to education and the estimation results provided in Table 1.4.2. The hypotheses for statistical significance of the effect of education on wages is given by:

$$H_0 : \beta_1 = 0$$

$$H_0 : \beta_1 \neq 0$$

From the estimation results, we notice that the t-ratio for the coefficient of years of schooling is 11.785. Note that the p-value is less than 0.05. Hence, using the p-value rule we will reject the null hypothesis and conclude that the effect of education on wages is **statistically significant**. You can also look at the critical value from the t-distribution at 5% level of significance, two-sided, and with $N - 2 = 998$ degrees of freedom. The corresponding value is 1.96 which is less than the absolute value of the t-ratio, leading us to reject the null hypothesis as well.

1.6.1 Confidence interval for regression coefficients

Note that for every regression coefficient in our model, we can compute the 95% confidence interval. For example, for the slope coefficient estimate, the 95% confidence interval is given by:

$$C.I.(\hat{\beta}_1) = \left[\hat{\beta}_1 - t_{c,2-sided} \times s.e.(\hat{\beta}_1), \hat{\beta}_1 + t_{c,2-sided} \times s.e.(\hat{\beta}_1)\right]$$

where where $t_{c,2-sided}$ is the critical value that can be obtained from the t-distribution table for a given level of significance (usually 5%) and degrees of freedom.

Example 1.3 (Confidence interval)**.** Again, using the returns to education example and results from Table 1.4.2, the 95% confidence interval is given by:

$$[0.106 - 1.96 \times 0.009, 0.106 + 1.96 \times 0.009] = [0.088, 0.124]$$

Notice that becuase, 0 is not part of the confidence interval, we can infer that the effect of education on wage is statistically significant.

Problems

Exercise 1.1. Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i^3 + \epsilon_i$$

a. Is this model considered a **linear regression** in parameters?

b. Write down the equation for the population regression function.

c. Suppose after using OLS you obtain: $\hat{\beta}_0 = 12$ and $\hat{\beta}_1 = -0.14$. Write down the equation for the predicted value.

Exercise 1.2. Suppose you have the following data on two variables:

ID	Y_i	X_i
1	7	4
2	9	8
3	11	5
4	13	7

a. Using OLS, compute the estimator for the slope coefficient (β_1) and the intercept ($ _0) for the following model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

b. Compute residuals (e_i) for each observation.

c. Show that $\sum_{i=1}^{4} e_i = 0$ and $\sum_{i=1}^{4} e_i X_i = 0$.

Exercise 1.3. Suppose you have the following data on sales (Y) and advertising expenditure (X):

ID	Y_i	X_i
1	8	5
2	10	9
3	12	7
4	14	9

a. Using OLS, compute the estimator for the slope coefficient (β_1) and the intercept ($ _0) for the following model:

$$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + \epsilon_i$$

b. Interpret the estimated slope coefficient.

Exercise 1.4. Suppose you have the following regression model:

$$Y_i = \beta_0 + \beta_1 ln(X_i) + \epsilon_i$$

After estimating this model in R, you obtain the following output:

Variable	b	s.e.(b)	t-stat	p-value
Intercept	-0.004	0.003	1.333	0.242
Ln(X)	0.887	0.277	3.202	0.003

Suppose sample size is 36 and the R-squared is 0.48.

a. Interpret the slope coefficients.

b. Test whether the slope coefficient is statistically significant.

c. Compute the 95% confidence interval for the slope coefficients.

d. Interpret R^2 of this model.

Solutions

Exercise 3.1:

a. Yes, because both β_0 and β_1 enter linearly in the model.

b. The population regression function is given by:

$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i^3$$

c. The predicted value is given by:

$$\widehat{Y_i} = 12 - 0.14X_i^3$$

Exercise 3.2:

a. Using OLS we get $\widehat{\beta_0} = 6.4$ and $\widehat{\beta_1} = 0.6$.

b. $e_i = Y_i - \widehat{Y_i} = Y_i - 6.4 - 0.6X_i$. Hence,

ID	Y_i	X_i	e_i
1	7	4	-1.8
2	9	8	-2.2
3	11	5	1.6

ID	Y_i	X_i	e_i
4	13	7	2.4

c. It is easy to show that sum of residuals from above table is 0. Similarly you can show the sum of residuals multiplied by X is 0.

Exercise 3.3:

a. You need to first compute natural logs of both Y and X. You will get:

ID	Y_i	X_i
1	2.08	1.61
2	2.30	2.20
3	2.48	1.95
4	2.64	2.20

Now apply OLS to get $\widehat{\beta_0} = 1.1191$ and $\widehat{\beta_1} = 0.6311$.

b. Because both variables are in logs, the interpretation of the slope is in terms of elasticity. Hence, if X increases by 1%, then Y increases by 0.6311%.

Exercise 3.4:

a. Because Y is in levels but X is in logs, the interpretation is in terms of semi-logs. So if X increases by 1 %, Y increases by 0.00887 units.

b. Using the p-value rule, the R output suggests that the coefficient of X is statisticall significant. This is because p-value is 0.003 which is less than the level of significance of 5%. Hence, we will reject the null hypothesis that the slope coefficient is equal to 0.

c. Using the confidence interval formula and critical value from t-distribution of 2.042 we get:

$$CI(\beta_1) : 0.887 \pm 0.277 \times 2.042 = (0.3214, 1.4526)$$

d. An R-squared of 0.48 indicates the 48% of the total variation in the dependent variable is explained by our model.

Chapter 2

Multiple regression model

In the previous chapter we modeled wages as a function of education only. However, it is reasonable to argue that there are other characteristics of individuals that can affect their wages. For example, industry of employment can affect wages with those in financial industry earning a higher wage on average than those engaged in retail industry. More importantly, there is an argument for this variable having an effect on wages which is independent of education. As a result, ignoring this variable will **confound** the effect of education on wages. In order to compute the pure effect of education on wages, we must compare those employed in the same industry but have different levels of education. This is what we mean by **controlling** for other factors in the regression and it is closely related to the concept of *ceteris paribus (all else equal)* in economics.

Note that ther can be two types of confounding factors:

1. Observed confounding factors: these, as the name suggests, are measurable relevant variables that can affect the outcome variable of interest. The idea here is that two observations may differ across obersvable dimensions which in turn influences the difference in their outcomes. This is known as **observed heterogeneity**. The multiple regression model can address this problem by adding all relevant and measurable independent variables to the right hand side of the regression model.

2. Unobserved confounding factors: these variables by definition are unobserved or difficult to measure in real world. The idea here is that two observations may differ across unobersvable dimensions which in turn influences the difference in their outcomes. This is known as **unobserved heterogeneity**. Going back to our wages regression model, it is possible that two inidividuals with same level of education and industry of employment earn different wages due to differences in their innate ability which is unobserved and very difficult to measure. There are some methods in Econometrics that address this problem and this problem posese one of the most serious challenge to the credibility of the estimated regression cofficients using the OLS. More on this later.

2.1 Multiple Regression Model

The multiple regression model simply adds more independent variables on the right-hand side of the regression model. Suppose Y_i denotes our outcome variable of interest for observation i and we believe there are K possible determinants of this outcome, denoted by $X_{i1}, X_{i2}, ..., XiK$. The population regression function (PRF), which is the conditional expectation of Y_i given data on all X variables is given by:

$$E(Y_i|X_{1i}, X_{2i}, ..., X_{Ki}) = f(X_{1i}, X_{2i}, ..., X_{Ki})$$

The resulting regression model is called **multiple regression** model:

$$Y_i = E(Y_i|X_{1i}, X_{2i}, ..., X_{Ki}) + \epsilon_i$$

Now we need to make an assumption about how each X variable affects the dependent variable. For example, if all effects are linear, then we get the following regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_K X_{Ki} + \epsilon_i$$

An important distinction from the simple regression model is the way we interpret the slope coefficients. Now, β_1 captures the effect of a unit change in X_1 on Y, **__holding all other X-variables constant__. In this sense, β_1 is the partial slope as it measures the pure effect of X_1 on Y after partialing or netting out the effects of all other included X-variables on Y.

For the above model, the predicted value of Y_i is given by:

$$\hat{Y_i} = \hat{\beta_0} + \hat{\beta_1} X_{1i} + \hat{\beta_2} X_{2i} + ... + \hat{\beta_K} X_{Ki}$$

and the residuals are given by:

$$e_i = Y_i - \hat{Y_i} = Y_i - (\hat{\beta_0} + \hat{\beta_1} X_{1i} + \hat{\beta_2} X_{2i} + ... + \hat{\beta_K} X_{Ki})$$

Note that the OLS minimization problem now involves choosing $\hat{\beta_0}, \hat{\beta_1}, \hat{\beta_2}, ..., \hat{\beta_K}$ that will minimize the sum of residuals squared (RSS) where

$$RSS = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} [Y_i - (\hat{\beta_0} + \hat{\beta_1} X_{1i} + \hat{\beta_2} X_{2i} + ... + \hat{\beta_K} X_{Ki})]^2$$

Example 2.1. Suppose we are interested in finding out what labor market characteristics affect wages. For this purpose, we collect data on wages, education and experience. Suppose the regression model is given by:

$$ln(Wage_i) = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + \epsilon_i$$

where $Wage_i$ is the annual wages and salaries of individual i in dollars. $Education_i$ denotes years of education of individual i. Finally, $Experience_i$ denotes years of experience.

Table 2.1 below presents the estimation results for the extended wage regression. Note that now the interpretation of the coefficient of education is as follows: holding the level of experience constant at its mean, raising education by 1 year will increase wages by 11.8%. Also notice that the coefficient of experience is positive indicating that more experienced workers earn higher wage on average.

OLS Estimation of Earnings Equation

Table 2.1: OLS Estimation of Earnings Function

	Dependent variable:	
	Log of Annual Wages	
	(1)	(2)
Intercept	9.259***	8.792***
	(0.135)	(0.150)
Education	0.106***	0.118***
	(0.009)	(0.009)
Experience		0.013***
		(0.002)
Observations	1,000	1,000
R^2	0.122	0.160
Adjusted R^2	0.121	0.158
Residual Std. Error	0.748 (df = 998)	0.732 (df = 997)
F Statistic	138.887*** (df = 1; 998)	94.937*** (df = 2; 997)
Note:		*p<0.1; **p<0.05; ***p<0.01

2.2 Goodness of fit redux

Table 2.1 shows results for two models: a simple regression model with education as the only explanatory model and a multiple regression model that adds experience to the simple model. A natural question to ask is whether adding these variables **matter** for explaining data on wages. As discussed in Chapter 3, R^2 provides one such measure. Higher value indicates greater variation of the dependent variable is explained by our included X-variables. However, we cannot compare these two models based on only R^2. The reason is by definition R^2 is guaranteed to increase when we add more independent variables to a model. This is true regardless of whether the variables we add our relevant to the problem at hand or not. In order to compare these two models with different number of explanatory variables, we need to acknowledge the following two consequences of adding an X-variable to our model:

i. The residual variation (unexplained) goes down improving the fit of the line. This is the **benefit** of adding an X-variable to the model.

ii. The degrees of freedom decreases by 1. This loss makes our estimators less precise and also lowers the power of any hypothesis test we may wish to conduct. This is the **cost** of adding an X-variable.

When we use R^2 to compare models with different number of explanatory variables, we are only considering the benefit of doing so without any regard to the cost. A more balanced measure would account for both the benefit in terms of improved fit and the cost in terms of the loss of the degrees of freedom. **Adjusted-R^2** is one such measure and is given by:

$$\text{Adjusted-}R^2 = 1 - \left(\frac{RSS}{TSS}\right) \times \left(\frac{N-1}{N-K}\right)$$

For a given sample size, as we add more X-variables, K increases. In the above formula, RSS goes down and hence the first term in the parenthesis given by $\left(\frac{RSS}{TSS}\right)$ becomes smaller. But N-K goes down as well and the second term in the parenthesis given by $\left(\frac{N-1}{N-K}\right)$ becomes smaller. The net effect on **Adjusted-R^2** depends on which of these two effects, capturing benefit and cost previously defined, is stronger. As a result, unlike R^2, this measure does not increase just by adding more X-variables to the model.

As a result, we can compare the two models presented in table using adjusted-R^2 with a greater value indicating better fit. From Table 2.1 we observe that adding experience to the model increase the fit from 12.1% to 15.8%, indicating that including experience allows us to explain greater amount of observed variation in wages in our sample.

2.3 Hypothesis testing in a multiple regression model

In addition to the test of statistical significance for each included X-variable in our model, we can now conduct different kinds of tests driven by economic theory. In this section we will use the production function and economic theory to illustrate hypothesis testing in a multiple regression framework.

Example 2.2 (Cobb-Douglas Production Function)**.** Consider the following Cobb-Douglas production function where output (Y) is a function of labor (L), capital (K) and material (M). In addition there is a constant that captures existing technology (A):

$$Y_i = A \times K_i^{\beta_1} \times L^{\beta_2} \times M^{\beta_3}$$

Note that the value of $\beta_1 + \beta_2 + \beta_3$ captures the **economies of scale** with three possibilities:

1. Constant returns to scale implied by $\beta_1 + \beta_2 + \beta_3 = 1$
2. Decreasing returns to scale implied by $\beta_1 + \beta_2 + \beta_3 < 1$

3. Increasing returns to scale implied by $\beta_1 + \beta_2 + \beta_3 > 1$

Using the multiple regression model, we can easily estimate the above production function and formally test what kind of returns to scale are prevalent in our sample. To do that first rewrite the above production function in natural logs as follows:

$$ln(Y_i) = ln(A) + \beta_1 ln(K_i) + \beta_2 ln(L_i) + \beta_3 ln(M_i)$$

This shows that a Cobb-Douglas production function can be approximated by a relationship which linear in natural logs of output, labor, captial, and material. The implied regression model is given by:

$$ln(Y_i) = \beta_0 + \beta_1 ln(K_i) + \beta_2 ln(L_i) + \beta_3 ln(M_i)\epsilon_i$$

We can estimate the production function either by using a cross-sectional data on firms at a point in time, or use data of one firm over time. In this example, we will use a cross-section of 50 firms to estimate the production function.

Table 2.3 provides the results of the estimation of the above model. From the p-values for each slope coefficient, we can infer that each effect is statistically significant at the 5% level of significance. In terms of economic significance, we find that 1% increase in capital increases output by 0.293%, holding constant labor and material inputs at their mean. In contrast, a 1% increase in labor increases output by 0.534%, holding constal capital and labor at their means. Finally, holding labor and capital constant, a 1% increase in material input raises output by 0.264%. Based on this we can argue that labor input has most economic significance. We also find that our model can explain 80% of the total variation in output observed in our sample as indicated by R^2 value. Next, we will conduct two types of hypotheses which are only possible in a multiple regression case.

OLS Estimation of Cobb-Douglas Production Function

2.3.1 Test of statistical significance of the entire model

Here our goal is to test whether all slope coefficients included in our model are jointly equal to 0. If true, then our model does not add any explanation for the output. Formally, in our example, we are testing:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_A : \text{ Not } H_0$$

Table 2.2: OLS Estimation of Cobb-Douglas Production Function

	Dependent variable:
	Natural Log of Output
Intercept	25.439^{***}
	(4.354)
Natural Log of Capital	0.293^{***}
	(0.051)
Natural Log of Labor	0.534^{***}
	(0.051)
Natural Log of Material	0.264^{***}
	(0.049)
Observations	50
R^2	0.800
Adjusted R^2	0.787
Residual Std. Error	10.485 (df = 46)
F Statistic	61.377^{***} (df = 3; 46)
Note:	*p<0.1; **p<0.05; ***p<0.01

One way to think of the above null hypothesis is that it places a **restriction** on our model. If this condition is true then our model can be reduced to:

$$ln(Y_i) = \beta_0 + \epsilon_i$$

The above model is called the **restricted** model and it will gives a value of R^2 which we will denote it R^2_R. Note that because there is no variable in our restricted model here, $R^2_R = 0$ by definition. Our original model with labor, capital, and material included also gives us an R^2 which we will denote by R^2_U. If the null hypothesis is true, then dropping labor and capital from the model should not lead to a **significant** decline in goodness of fit (or R^2), i.e, the difference between R^2_U and R^2_R should not be large. Our test statistic for this test is given by:

$$F = \frac{(R^2_U - R^2_R)/J}{(1 - R^2_U)/(N - K - 1)}$$

where J denotes the number of restrictions we impose in our null hypothesis and K denotes number of X-variable in the original model. In this example both J and K take the value of 3. Under the null hypothesis, this F statistic follows F-distribution with J degrees of freedom for the numerator and $N - K - 1$ degrees of freedom for the denominator. We can use the F-distribution table to get the critical value denoted by F_c. The decision rule is if $F > F_c$ then reject the null hypothesis.

In our example, $R^2_U = 0.719$. Also note that in this case, $R^2_R = 0$. Hence, the F statistic value is:

$$F = \frac{(0.80 - 0)/3}{(1 - 0.80)/(50 - 3 - 1)} = 61.33$$

The critical value at 5% level of significance can be obtained from the F-distribution table using $df_1 = J = 3$ and $df_2 = N - K - 1 = 46$. In our case this value is 2.81. Because the F statistic is greater than the critical value, we reject the null hypothesis and conclude that our model is able to explain a statistically significant amount of variation in output.

2.3.2 Testing whether one or more of the variables can be eliminated from the model

In the previous example, we imposed the restriction that all three inputs have zero effect on output. However, often we maybe interested in finding out whether any one variable or a subset of included X-variables in the model can be dropped without significantly impacting the fit of the model. For example, suppose we

want to test whether dropping capital and material inputs from our model leads to a significant loss in the fit of the model. Here, our hypotheses are:

$$H_0 : \beta_1 = \beta_3 = 0$$
$$H_A : \text{Not } H_0$$

Our restricted model, if the statement in the H_0 is true is given by:

$$ln(Y_i) = \beta_0 + \beta_2 ln(L_i) + \epsilon_i$$

Note that if we estimate the above restricted model, we will obtain an R^2 value which will be non-zero but lower than the original model with capital and material inputs added. We can test whether the reduction in R^2 is significant or not. Table 2.3.2 below presents the results of for the restricted model. For convenience, I also add the original unrestricted model results from Table 2.3 as well. We note that $R^2_U = 0.80$ and $R^2_R = 0.510$. We can compute the F statistic as follows:

$$F = \frac{(0.80 - 0.510)/2}{(1 - 0.80)/(50 - 3 - 1)} = 33.35$$

Note how in this example, we use $J = 2$ and $K = 3$ in the F statistic formula. This is because we are now dropping two variables from our orignal model implying that we are imposing two restrictions and hence $J = 2$.}

The critical value at 5% level of significance and $df_1 = 2, df_2 = 46$ is 3.20. Because the F statistic is greater than the critical value, we reject the null hypothesis and conclude that capital and material input add significantly to explanation of output variation in our sample.

OLS Estimation of Cobb-Douglas Production Function

2.3.3 Testing a linear restriction on slope coefficients

Finally, we can also use our multiple regression model to test a linear restriction on the slope coefficients. Often such restrictions are informed by economic theory and vary from applications to applications. In our example, one such restriction is the returns to scale. Formally in our original model we can test the following hypotheses,

$$H_0 : \beta_1 + \beta_2 + \beta_3 = 1$$

Table 2.3: OLS Estimation of Cobb-Douglas Function

	Dependent variable:	
	Natural Log of Output	
	(1)	(2)
Intercept	25.439***	52.819***
	(4.354)	(4.265)
Natural Log of Capital	0.293***	
	(0.051)	
Natural Log of Labor	0.534***	0.550***
	(0.051)	(0.078)
Natural log of Material	0.264***	
	(0.049)	
Observations	50	50
R^2	0.800	0.510
Adjusted R^2	0.787	0.499
Residual Std. Error	10.485 (df = 46)	16.077 (df = 48)
F Statistic	61.377*** (df = 3; 46)	49.887*** (df = 1; 48)
Note:		*p<0.1; **p<0.05; ***p<0.01

$$H_A : \beta_1 + \beta_2 + \beta_3 \neq 1$$

If the null hypothesis is true, then we have constant returns to scale. Also notice that we are using a two-sided alternative hypothesis here. As result, rejection of the null hypothesis will imply non-constant returns to scale. But we cannot infer whether we have decreasing or increasing returns. For that we need to specify an appropriate one-sided alternative. For example, for decreasing returns to scale as our alternative, we would specify:

$$H_A : \beta_1 + \beta_2 + \beta_3 < 1$$

In this example, we will be using a two-sided alternative. There are two ways of conducting this kind of a hypothesis test.

- **Method 1: t-test**

Here, we can compute the t-ratio given by:

$$t = \frac{(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3) - 1}{s.e.(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3)}$$

Note that Table 2.3 gives us the three estimated regression coefficients. But in order to compute the standard error we also need pairwise correlations between estimated regression coefficients. This is because by definition:

$$s.e.(\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3) = \sqrt{Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + Var(\hat{\beta}_3) + 2Cov(\hat{\beta}_1, \hat{\beta}_2) + 2Cov(\hat{\beta}_1, \hat{\beta}_3) + 2Cov(\hat{\beta}_2, \hat{\beta}_3)}$$

The covariance matrix for 3 regression coefficients is presented in Table 2.4. The diagonal elements of this matrix gives us the variance of each coefficient whereas the off diagonal elements gives us the covariance between a pair of the coefficients. We can use that information and compute the standard error as follows:

$$s.e.(\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3) = \sqrt{0.0026 + 0.0026 + 0.0024 + 2 \times (-0.0001) + 2 \times (-0.002) + 2 \times 0.000} = 0.0584$$

Then, the t-statistic is given by:

$$t = \frac{(0.293 + 0.534 + 0.264) - 1}{0.0584} = 1.56$$

Table 2.4: Variance-Covariance Matrix

	$\hat{\beta_0}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{\beta_3}$
$\hat{\beta_0}$	18.9554	-0.1220	-0.1130	-0.1060
$\hat{\beta_1}$	-0.1220	0.0026	-0.0001	-0.0002
$\hat{\beta_2}$	-0.1130	-0.0001	0.0026	0.0000
$\hat{\beta_3}$	-0.1060	-0.0002	0.0000	0.0024

Under the null hypothesis, our test statistic follows t-distribution with $N-K-1 = 46$ degrees of freedom. The critical value at 5% level of significance for a two-sided alternative is 2.009. Because |t| is less than the critical value, we do not reject the null hypothesis. Hence, in our sample we do not find evidence against the constant returns to scale assumption.

- **Method 2: F-test**

An alternative way to test a linear restriction is to follow the same method we used for previous two tests. We can use the null hypothesis restriction and rewrite our original model as:

$$ln(Y_i) = \beta_0 + (1 - \beta_2 - \beta_3)ln(K_i) + \beta_2 ln(L_i) + \beta_3 ln(M_i)\epsilon_i$$

or equivalently,

$$ln(Y_i) - ln(K_i) = \beta_0 + \beta_2[ln(L_i) - ln(K_i)] + \beta_3[ln(M_i) - ln(K_i)] + \epsilon_i$$

Note here we have replaced $\beta_1 = 1-\beta_2-\beta_3$ using the linear restriction imposed by the null hypothesis. Consequently, the model specification written above is our restricted model and we can use the F-test to investigate whether imposing this restriction leads to a significant reduction in the model fit. The corresponding F-statistic in this case is given by 0.957. The critical value with $df_1 = 1$ and $df_2 = 46$ is 4.05. Because the F statistic is less than the critical value, we do not reject the null hypothesis.

Problems

Exercise 4.1. In a multiple regression analysis, a model has three independent variables. The analyst decides to add another (fourth) independent variable while retaining the other three independent variables. What will happen to R^2 due to this addition? Explain.

Exercise 4.2. Suppose you have the following estimated model:

Table 2.5: OLS Estimation of Cobb-Douglas Function

	Dependent variable:	
	Natural Log of Output	
	(1)	(2)
Intercept	15.431***	15.013***
	(4.289)	(4.025)
Natural Log of Income	0.673***	0.675***
	(0.094)	(0.093)
Natural Log of Wealth	0.398***	0.393***
	(0.056)	(0.053)
Interest (	(0.188)	
Observations	40	40
R^2	0.729	0.729
Adjusted R^2	0.707	0.714
Residual Std. Error	6.846 (df = 36)	6.762 (df = 37)
F Statistic	32.347*** (df = 3; 36)	49.683*** (df = 2; 37)
Note:		*p<0.1; **p<0.05; ***p<0.01

Solutions

Exercise 4.1. As discussed in section 4.2, mathematically RSS of the bigger model (with 4 variables) has to be lower than the RSS of the smaller model (with 3 variables). This is because OLS chooses slope coefficients by minimizing RSS. Now as we know that

$$R^2 = 1 - \frac{RSS}{TSS}$$

, mechanically a lower RSS value implies a larger R^2. So in our example, an additional variable added to the model will lead to an increase in R-squared regardless of whether the additional variable is relevant to the model or not.

Chapter 3

Functional form and dummy variables

Suppose we are interested in estimating a relationship between Y and X. In general the population regression function is an unknown function of X, that is:

$$E(Y_i|X_i) = f(X_i)$$

Our discussion on regression model thus far has mostly focused on the linear relationship between Y and X variable, with the exception of using natural log transformations. In this chapter we will look at two alternative ways of capturing non-linearity in economic relationships. Our objective is to find specifications for $f(X_i)$ that allow for non-linear terms in X.

3.1 Polynomials in the regression model

One way to capture non-linear relationship between the dependent variable, Y, and the independent variable, X is to use polynomials in X. For example, consider the following polyonomial of order 2 (quadratic):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

In the above model, the first and second derivatives of Y with respect to X allow us to capture the non-linear relationship between these two variables. The first derivative of Y with respect to X gives us:

$$\frac{dY_i}{dX_i} = \beta_1 + 2\beta_2 X_i$$

Note that:

1. The effect of X on Y now depends on the level of X.
2. We cannot interepret β_1 as the effect of X_i on Y, holding X_i^2 constant. This is because when X_i changes, X_i^2 changes as well. Hence, best way is to simply plot the non-linear relationship and/or compute the effect on predicted value of Y due to change in X.
3. Further, taking the second derivative we get:

$$\frac{d^2Y_i}{dX_i^2} = 2\beta_2$$

Hence, the sign of β_2 tells us whether the relationship is *concave* ($\beta_2 < 0$) or *convex* ($\beta_2 > 0$).

4. Finally, we can set the first derivative equal to 0 and solve for X^* as follows:

$$\frac{dY_i}{dX_i} = \beta_1 + 2\beta_2 X_i = 0 \Rightarrow X_i^* = -\frac{\beta_1}{2\beta_2}$$

X_i^* will either maximize or minimize Y depending on the sign of β_2.

3.1.1 Obtaining optimal polynomial order using goodness of fit

In general we do not know the order of polynomial and in theory can express the relationship between Y and X as a polynomial of order q given by:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + ... + \beta_q X_i^q + \epsilon_i$$

In order to determine what value of q best explain our data we need to compare models with different number of explanatory variables. For example, to compare a quadratic model with a cubic model, we are comparing a model with two explanatory variables (X_i and X_i^2) with a model with three independent variables (X_i, X_i^2, and X_i^3). Such comparisons require us to incorporate the

tradeoff between improved fit and loss of degrees of freedom. In practice, do determine a value for q we can follow the following procedure:

Step 1: Pick a maximum value for q. This part is arbitrary and usually will affect your final answer. Let us denote this value as q_{max}.

Step 2: Estimate the polynomial of order q_{max}:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + ... + \beta_q X_i^{q_{max}} + \epsilon_i$$

This is our unrestricted model and we denote R-squared from this model as R_U^2.

Step 3: Estimate the polynomial of order $q_{max} - 1$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + ... + \beta_q X_i^{q_{max}-1} + \epsilon_i$$

This is our restricted model and we denote R-squared from this model as R_R^2.

Step 4: Test whether there is a loss in fit between step 2 and step 3 using the F-test:

$$H_0 : \text{True model is polynomial of order } q_{max} - 1$$
$$H_A : \text{True model is polynomial of order } q_{max}$$

The test statistic is given by:

$$F = \frac{(R_U^2 - R_R^2)/1}{(1 - R_U^2)/(N - q_{max} - 1)}$$

Under the null-hypothesis, the test statistic follows F distribution with $v_1 = 1$ and $v_2 = N - q_{max} - 1$ degrees of freedom. If you reject the null hypothesis, stop and conclude that $q = q_{max}$ is the final model.

Step 5: If you do not reject the null hypothesis in Step 4, continue the process until you find the highest order polynomial where you reject the null hypothesis.

Example 3.1. Suppose we are interested in estimating the relationship between wages (Y) and labor market characteristics of workers. Specifically, we collect data on years of education (X_1) and years of experience (X_2). Consider the following specification:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \epsilon_i$$

In the above model we assume that the relationship between wages and years of experience is quadratic. We can test this formally by estimating the following restricted version of the model with only linear term in experience.

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

If the true relationship between experience and wages is indeed quadratic, then estimating the restricted model will result in substantial loss of fit and hence will be rejected by our F-test. Table 3.1.1 below provides the estimation results of these two models. Notice that there is an drop in R^2 from 19.8%, to 16% indicating that removal of squared experience does lead to loss of fit. Next, we test whether this loss in fit is statistically significant using the F-test:

$$H_0 : \text{True model is linear in experience}$$

$$H_A : \text{True model is quadratic in experience}$$

$$F = \frac{(R^2_U - R^2_R)/1}{(1 - R^2_U)/(N - 3 - 1)} = \frac{0.198 - 0.16}{(1 - 0.198)/(100 - 3 - 1)} = 47.19202$$

The critical value from F-table with $v_1 = 1$ and $v_2 = 1000 - 3 - 1 = 996$ is 3.84. Because the F statistic is greater than this critical value, we reject the null hypothesis and conclude that the relationship between wages and experience is quadratic. Also note that the coefficient of the squared term in experience is negative implying a concave relationship between experience and wages. Finally, we can also compute the level of experience that maximizes wages by taking the first derivative with respect to experience and setttin it 0:

$$\frac{dY_i}{dX_{2i}} = \beta_2 + 2\beta_3 X_{2i} = 0 \Rightarrow X^*_{2i} = -\frac{\beta_1}{2\beta_2}$$

Given our estimates, this indicates that for $X^*_{2i} = -\frac{0.058}{2 \times (-0.001)} = 29$ years, we expect wages to reach a maximum with respect to experience and then level off or decline after that.

The use of the squared experience term is motivated by the empirical literature on determinants of earnings. As discussed earlier, the orginal Mincerian equation included a squared experience to capture the non-linear relationship between wages and experience.

OLS Estimation of Earnings Equation

Table 3.1: OLS Estimation of Earnings Function

	Dependent variable:	
	Natural Log of Wages	
	(1)	(2)
Intercept	8.455^{***}	8.792^{***}
	(0.154)	(0.150)
Education	0.113^{***}	0.118^{***}
	(0.009)	(0.009)
Experience	0.058^{***}	0.013^{***}
	(0.007)	(0.002)
Experience-squared	-0.001^{***}	
	(0.0001)	
Observations	1,000	1,000
R^2	0.198	0.160
Adjusted R^2	0.195	0.158
Residual Std. Error	0.716 (df = 996)	0.732 (df = 997)
F Statistic	81.916^{***} (df = 3; 996)	94.937^{***} (df = 2; 997)
Note:		*p<0.1; **p<0.05; ***p<0.01

3.2 Step functions: Dummy variables in the regression model

In many practical applications we may wish to estimate the effect of qualitative variables on an outcome of interest. For example,

a. what is the effect of a worker's occupation or race on wages?

b. do we have an increase in consumer spending during Christmas season?

c. how does college major affects starting salary of graduates? Or

In each of the above examples, we want to estimate the effect of **group membership** on the outcome of interest. For example, do observations that belong to Christmas season have higher consumer spending than observations in non-Christmas season. One important thing to note here is that incorporating such variables in our regression mean that we need a **step-function** because we beleive there is a discrete jump in the outcome variable as we compare groups.

By definition qualitative variables are non-numeric in nature. One way to incorporte such variables is to use **dummy variables**. A dummy variable is an indicator variable that takes value of 1 if an observation belongs to a particular category and 0 otherwise. For example, suppose we are interested in finding the effect of race on wages. Then, we can define a dummy variable, D_i as follows:

$$D_i = \begin{cases} 1 & \text{if observation } i \text{ is white} \\ 0 & \text{otherwise} \end{cases}$$

Once we have defined the dummy variable, we can add it to the regression model as any other independent variable. Note that:

1. A qualitative variable with J categories can be incorporated by adding $J-1$ dummy variables in the regression model. The excluded category is called the **base group** and serves as the benchmark category relative to which the effect on the dependent variable is measured. For example, in the dummy variable we defined above, we have two categories in our data: white and non-white. The way we have defined the dummy variable implies that our excluded group is non-white workers. This is to avoid the *dummy variable trap*, where adding J dummy variables for a qualitative variable with J categories leads to perfectly linear relationship between included independent variables in our model.

2. The effect on the regression model depends on how we add the dummy variable to our regression model. There are two options:

2.1. Intercept dummy: here we are only interested in the categorical effect alone. This can be done by adding a dummy variable to the regression model and has the effect of shifting the intercept of the regression model.

2.2. Slope dummy: here we are interested in capturing any interaction a categorical variable may have with other indepedent variables in the regression model. This can be done by adding a product of the dummy variable with the independent variable in question and has the effect of shifting the slope of the regression model.

Example 3.2 (Intercept Dummy in Earnings Function)**.** Let us continue our example of the earnings function. Now suppose we want to find out whether being male has an effect on wages,i.e, if we compare two workers with same years of education (X_1) and experience (X_2), but one is male and other is female, do we see any difference in their wages. This can be done by estimating the following regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \beta_4 Male_i + \epsilon_i$$

where,

$$Male_i = \begin{cases} 1 & \text{if observation } i \text{ is male} \\ 0 & \text{otherwise} \end{cases}$$

In this model, β_4 measures the effect of being male on wages for individuals with same years of education and experience. To see this, lets write down the predicted wage from our model for male and female workers separately:

$$\widehat{Y_i^m} = \widehat{\beta_0} + \widehat{\beta_1} X_{1i}^m + \widehat{\beta_2} X_{2i}^m + \widehat{\beta_3} (X_{2i}^m)^2 + \widehat{\beta_4}$$

$$\widehat{Y_i^f} = \widehat{\beta_0} + \widehat{\beta_1} X_{1i}^f + \widehat{\beta_2} X_{2i}^f + \widehat{\beta_3} (X_{2i}^f)^2$$

Here the superscripts m denotes male and f denotes female. Note that if we equalize education and experience between men and women, then:

$$\widehat{Y_i^m} - \widehat{Y_i^f} = \widehat{\beta_4}$$

Consquently, adding an intercept dummy simply changes the intercept of the regression line by the parameter of the intecept dummy. Table 3.2 provides the estimation results for this model. Note that the estimated coefficient of Male dummy variable is 0.306. This means that a male worker's wages is 30.6% higher than a female worker with same years of education and experience.

Earnings Equation with intercept dummy variable

Table 3.2: Earnings Equation with intercept dummy variable

	Dependent variable:	
	Natural Log of Wages	
	(1)	(2)
Intercept	8.455^{***}	8.792^{***}
	(0.154)	(0.150)
Education	0.113^{***}	0.118^{***}
	(0.009)	(0.009)
Experience	0.058^{***}	0.013^{***}
	(0.007)	(0.002)
Experience-squared	-0.001^{***}	
	(0.0001)	
Observations	1,000	1,000
R^2	0.198	0.160
Adjusted R^2	0.195	0.158
Residual Std. Error	0.716 (df = 996)	0.732 (df = 997)
F Statistic	81.916^{***} (df = 3; 996)	94.937^{***} (df = 2; 997)
Note:		*p<0.1; **p<0.05; ***p<0.01

Example 3.3 (Slope Dummy in Earnings Function)**.** In the previous example, we investigated whether being Male affects wages once we have controlled for education and experience. But in that analysis we assumed that there is not interaction between whether a worker is make with his years of education or experience. But is quite easy to think of interactions between the sex of a worker and their education or experience. For example, on average women tend to accumulate less years of experience than men of same age due to women taking more breaks from work due to child birth. Similarly, in many economies, men tend to have greater level of education, especially in emerging economies where parents allocate resources differentially between male and female child. To capture such interactions, we can use slope dummy variables. Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \beta_4 Male_i + \beta_5 (Male_i \times X_{1i}) + \epsilon_i$$

In this model, β_5 measures the differential effect education may have on wages depending on the sex of the worker. To see this note that:

$$\frac{\partial Y_i}{\partial X_{1i}} = \beta_1 + \beta_5 Male_i$$

As a result, for male workers, the effect of education on wages is $\beta_1 + \beta_5$ whereas for female workers this effect is β_1. Table 3.2 provides the estimation results for this model. Note that the estimated coefficient of the interaction between Male and education is positive but statistically insignificant. Hence, in our sample, there is no evidence that sex of workers interacts with their education.

Earnings Equation with intercept and slope dummy variables

Problems

Solutions

Table 3.3: Earnings Equation with intercept and slope dummy variables

	Dependent variable:	
	Natural Log of Wages	
	(1)	(2)
Intercept	8.455^{***}	8.792^{***}
	(0.154)	(0.150)
Education	0.113^{***}	0.118^{***}
	(0.009)	(0.009)
Experience	0.058^{***}	0.013^{***}
	(0.007)	(0.002)
Experience-squared	-0.001^{***}	
	(0.0001)	
Observations	1,000	1,000
R^2	0.198	0.160
Adjusted R^2	0.195	0.158
Residual Std. Error	0.716 (df = 996)	0.732 (df = 997)
F Statistic	81.916^{***} (df = 3; 996)	94.937^{***} (df = 2; 997)
Note:		*p<0.1; **p<0.05; ***p<0.01

Chapter 4

Classical assumptions and OLS estimator

4.1 Classical Assumptions

Similar to our discussion in chapter 2, a desirable sample estimator must be unbiased and efficient. The discussion so far has focused on estimating regression parameters using sample data on the dependent and the independent variables. We will now focus on the conditions under which the OLS estimator of regression parameters are **unbiased** and **efficient**.

Suppose we have the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

In the above model, β_0 and β_1 denote true but unknown population parameters of interest. We can use a sample data for Y_i and X_i to compute sample estimators for these two parameters. OLS is one such method of obtaining sample estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$. For these OLS estimators to be unbiased and efficient we need:

1. **Unbiasedness:**

$$E(\widehat{\beta_0}) = \beta_0 \text{ and } E(\widehat{\beta_1}) = \beta_1$$

2. **Efficiency**:

$$Var(\hat{\beta}_0) \text{ and } Var(\hat{\beta}_1) \text{ are smallest possible.}$$

In order for the OLS estimators to be unbiased and efficient, we need a set of assumptions to be satisfied. These assumptions are known as **classical assumptions** and the result is formally known as the **Gauss-Markov** theorem named after two mathematicians, namely, Carl Gauss and Andrey Markov.

Theorem 4.1 (Gauss-Markov theorem)**.** *The OLS estimator is the best linear and unbiased estimator (BLUE) if and only if the following six classical assumptions are satisfied:*

1. *The regression model is linear in parameters.*
2. *There is no linear relationship between included independent variables in the regression model or there is* ***no perfect multicollinearity****.*
3. *The expected value of the regression error term is 0.*

$$E(\epsilon_i) = 0 \text{ for all } i$$

4. *There is no heteroscedasticity, i.e, the variance of the regression error term is constant.*

$$Var(\epsilon_i) = \sigma^2 \text{ for all } i$$

5. *There is no serial correlation. In time series data serial correlation implies observations of a variable are correlated over time. This is also known as* ***auto-correlation****. One of the classical assumption is that there is no serial (or auto) correlation in regression error terms.*

$$Cor(\epsilon_t, \epsilon_{t-s}) = 0 \text{ where } t \neq s$$

Note that for cross-sectional data correlation in error terms across observations is known as **spatial correlation**. In this course we will abstract away from this type of correlation and focus only on the serial correlation in time series data.

6. No endogeneity problem, i.e, all included independent variables in the model are exogenous and hence are uncorrelated with the regression error terms.

$$Cor(X_{ki}, \epsilon_i) = 0 \text{ for} k = 1, 2, 3, .., K$$

Of these 6 assumptions, in practice, we often take assumptions 1 through 3 for granted and do not consider violations of these assumptions in our data. However, assumptions 3 to 6 are often not met in data and are investigated much more rigorously. Accordingly, in this chapter we will focus on heteroscedasticity, serial correlation, and no endogeneity.

4.2 Heteroscedasticity

One of the main assumptions that affect the efficiency of the OLS estimator is the assumption of no heteroscedasticity which forces a constant variance for the error term in our regression model. Consider the following simple regression model that explains food expenditure (Y) based on a person's disposable income (X):

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

The classical assumption of no heteroscedasticity (or Homoscedasticity) implies that $Var(\epsilon_i) = \sigma_\epsilon^2$. In the context of our example, this assumption can be interpreted as follows. The information a person's income has for his food expenditure does not vary by the any characteristic of this person (say income). As a result the range of errors we can make in predicting someone's food expenditure based on their income stays constant. Graphically, if error term are homoscedastic then there is no relationship between the range of errors we can make and a person's income. Graphically, Figure 6.1 below shows this pattern– whether we look at observations with low income or those with high income, the range of errors we make is more or less constant.

However, it is reasonable to argue that the value of information differ across observations in the following sense. Some observations have a greater role in reducing error variance than others. In our example, one can argue that food expenditure forms a bigger percent of a person with lower income than one with higher income who may use his income for non-food expenditure activities. In this case we would expect that the variance of errors will increase as income increases. This particular pattern of heteroscedastic errors is shown in Figure 6.2 below.

In general the exact form of heteroscedasticity will depend on the nature of the problem at hand. However, whether or not heteroscedasticity is present in our data is an empirical question.

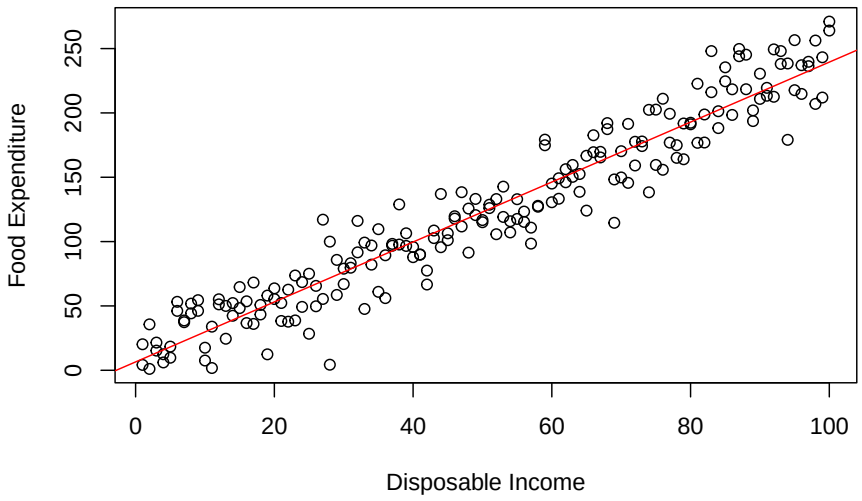


Figure 4.1: Homoscedastic Errors

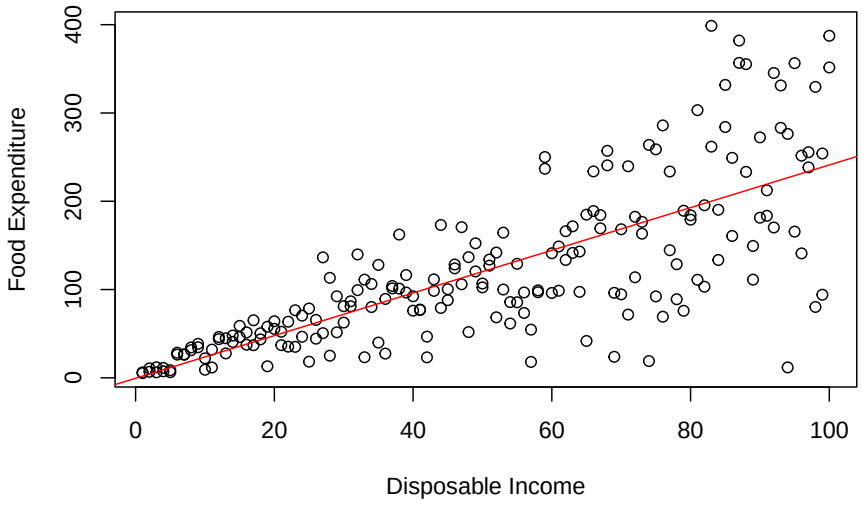


Figure 4.2: Heteroscedastic Errors

4.2.1 Consequences of Heteroscedasticity for the OLS estimator

Presence of heteroscedasticity is a violation of the classical assumption and accordingly will affect the properties of the OLS estimator. In general, if there is heteroscedasticity in data then:

1. The OLS estimator of each β coefficient is still unbiased.
2. However, due to ignoring the systematic variation in the error term, the OLS estimator is no longer efficient and the sample estimators of the standard errors of each β is incorrect. Consequently, we cannot conduct hypothesis testing on regression coefficients using the OLS estimator.

4.3 Testing for Hetroscedasticity in data

The first step for incorporating heteroscedasticity in our estimation is to test for its presence in our sample data. This test is based on OLS residuals of the original regression model and accounts for both linear and non-linear forms of heteroscedasticity. Consider the following regression model with two independent variables:

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

Using OLS we can obtain the residuals from this model, denoted by e_i. Next, we will use this residual data to test for the presence of heteroscedasticity.

4.3.1 LM test for linear heteroscedasticity: BP test

The first test of heteroscedasticity is the Breusch-Pagan (BP) test of linear heteroscedasticity. The procedure for implementing this test is detailed below:

Step 1. Estimate the regression model using OLS and obtain residuals: e_i

Step 2: Estimate the BP regression model where the dependent variable is squared residuals and all independent variables are included on the right hand side:

$$e_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \epsilon_i$$

Obtain R^2 from this regression and denote it by R^2_{BP}.

Step 3: The test of linear Heteroscedasticity is given by-

$$H_0 : \alpha_1 = \alpha_2 = 0 \Rightarrow \text{No linear heteroscedasticity}$$
$$H_A : Not\ H_0 \Rightarrow \text{linear heteroscedasticity}$$

The test statistic id denoted by LM and the formula is given by:

$$LM = R^2_{BP} \times N$$

where N denotes sample size. Under the null hypothesis this test statistic follows Chi-square distribution with J degrees of freedom, where J denotes number of independent variables in the BPG regression of Step 2. If the LM test statistic is greater than the critical value from the Chi-square distribution, we reject the null hypothesis and conclude that there is sample evidence for linear heteroscedasticity.

4.3.2 LM test for linear heteroscedasticity: White's test

We can also test for the presence of non-linear heteroscedasticity using White's test. The procedure for implementing this test is detailed below:

Step 1. Estimate the regression model using OLS and obtain residuals: e_i

Step 2: Estimate the BPG regression model where the dependent variable is squared residuals. Now, independent variables include all independent variables, squared of all independent variables, and their product. So for example with 2 independent variables X_1 and X_2 we get the following White regression:

$$e_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 (X_{1i} \times X_{2i}) + \epsilon_i$$

Obtain R^2 from this regression and denote it by R^2_{White}.

Step 3: The test of Hetroscedasticity is given by-

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0 \Rightarrow \text{No heteroscedasticity}$$
$$H_A : Not\ H_0 \Rightarrow \text{heteroscedasticity}$$

The test statistic id denoted by LM and the formula is given by:

$$LM = R^2_{White} \times N$$

where N denotes sample size. Under the null hypothesis this test statistic follows Chi-square distribution with J degrees of freedom, where J denotes number of independent variables in the White regression of Step 2. If the LM test statistic is greater than the critical value from the Chi-square distribution, we reject the null hypothesis and conclude that there is sample evidence for heteroscedasticity.

Example 4.1 (Testing for Hetroscedasticity)**.** One of the most important models in finance is the Fama-French 3-factor model of risk premium of a stock. As per this model, the expected return on an asset over and above a risk free rate depends on depends on three factors:

a. Market risk: Market return minus risk free rate

b. Size premium: small market captilization stocks tend to out perform large market capitalization stocks. This variable captures this permium.

c. Value premium: Stocks with high book-to-market value out perform those with low value. This variable in this sense captures the value premium.

The regression model implied is given by:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \epsilon_t$$

Here, y_t is the return on a stock minus the risk free rate. In this example we will use IBM stock and proxy risk free rate by using return on 1 month TB. X_{1t} denotes market risk, X_{2t} denotes size premium, and X_{3t} denotes value risk. The data for these three factors is downloaded from the following website:

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

In this application we use monthly data from Jan-2007 through June 2019. Table 4.3.2 provides the estimation of this model using OLS.

OLS Estimation of Fama-French 3-factor model

In order to use the OLS estimator for hypothesis testing, we need to confirm whether there is heteroscedasticity in our data. For this example the BP test regression is given by:

$$e_i^2 = \alpha_0 + \alpha_1 f_{1i} + \alpha_2 f_{2i} + \alpha_3 f_{3i} + \epsilon_i$$

The null hypothesis for no heteroscedasticity requires $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

The White test regression is given by:

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

The null hypothesis for no heteroscedasticity requires $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = 0$. We conduct both BP and White's test in R and present the results in Table 4.3.2. We find that there is evidence of heteroscedasticity according to both tests, as we reject the null hypothesis for

Table 4.1: OLS Estimation of Fama-French 3-factor model

	Dependent variable:
	Return on IBM-Risk Free Rate
Intercept	-0.054^{***}
	(0.010)
Market Risk Premium	0.014^{***}
	(0.002)
Size Premium	0.006
	(0.007)
Volume Premium	-0.012^{*}
	(0.006)
Observations	150
R^2	0.215
Adjusted R^2	0.199
Residual Std. Error	0.125 (df = 146)
F Statistic	13.330^{***} (df = 3; 146)
Note:	*p<0.1; **p<0.05; ***p<0.01

BP test at 5% level of significance and we reject the null hypothesis for White's test at 10% level of significance.

	BP test	White's test
LM statistic	10.099	15.396
p-value	0.018	0.081

Two tests of heteroscedasticity

4.4 Heteroscedasticity robust standard errors

In case we find heteroscedasticity in our data using OLS would lead to unbiased but inefficient estimators. The main consequence of heteroscedasticity, as discussed above, is that the OLS standard errors for each β coefficient in our regression model are incorrect. To see understand this issue consider the following simple regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Using OLS we obtain the following formula for the slope estimator:

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N}(X_i - \overline{X})^2}$$

With a little algebra, we can rewrite the above formula as follows:

$$\widehat{\beta_1} - \beta_1 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})\epsilon_i}{\sum_{i=1}^{N}(X_i - \overline{X})^2}$$

Now, by definition, $Var(\widehat{\beta_1}) = E(\widehat{\beta_1} - \beta_1)^2$. Hence, using above equation we get the formula for the variance of $Var(\widehat{\beta_1})$. Below we first compute this formula without making an assumption of no heteroscedasticity:

$$Var(\widehat{\beta_1}) = \frac{\sum_{i=1}(X_i - \overline{X})^2 E(\epsilon_i^2)}{\sum_{i=1}^{N}((X_i - \overline{X})^2)^2} \tag{4.1}$$

Note that in the above variance formula, $E(\epsilon_i^2) = Var(\epsilon_i)$. Equation (4.1) is the correct formula for computing the variance (and hence the correct standard error for $\widehat{\beta_1}$) when there is heteroscedasticity in data.

In contrast, OLS assumes homoscedasticity, implying $E(\epsilon_i^2) = Var(\epsilon_i) = \sigma_\epsilon^2$ (a constant). As a result the OLS estimator for the variance of $Var(\widehat{\beta_1})$ is given by:

$$Var(\widehat{\beta_1}) = \frac{\sum_{i=1}(X_i - \overline{X})^2\sigma_\epsilon^2}{\sum_{i=1}^{N}(X_i - \overline{X})^2} = \frac{\sigma_\epsilon^2}{\sum_{i=1}^{N}(X_i - \overline{X})^2} \tag{4.2}$$

Note that the OLS variance formula in is different from the formula for variance in equation (4.1).

In applied economic analysis it is common to replace the incorrect OLS standard errors (square root of equation (4.2)) with the **hetroscedasticity robust standard errors** (square root of equation (4.1)). But for that we need to estimate $E(\epsilon_i^2)$ using our sample data. One of the most commonly used correction was proposed by the economist Hal White in 1980. He suggested replacing $E(\epsilon_i^2)$ in equation (1) with e_i^2 where e_i is the OLS residual. Hence, the formula for White's corrected standard error is given by:

$$Var(\widehat{\beta_1}) = \frac{\sum_{i=1}(X_i - \overline{X})^2 e_i^2}{\sum_{i=1}^{N}((X_i - \overline{X})^2))^2}$$

Table 4.4 reports the White's standard errors for the regression model presented in Table 4.3.2. You can see minor differences in the standard errors for each β coefficient after correcting for heteroscedasticity.

	White's Robust Standard Errors
(Intercept)	0.011
Market Risk Premium	0.003
Size Premium	0.007
Volume Premium	0.005

White's Standard Errors

Note that use of robust standard errors may mask a more serious issue with your regression model. For example, there may be important differences between subgroups of observation in your sample. Similarly, the true relationship may not be linear. Such misspecification issues can manifest themselves as heteroscedasticity. If the robust standard errors using White's method are very different from the OLS standard errors then one may have to worry about the misspecification issues and simply using robust standard errors is not a good idea.

4.5 Serial correlation

In time series data it is common to observe correlation across observations over time. Indeed many relationships in economics are dynamic in nature. For

example, consumption habits indicate past consumption has an effect on current consumption. Similarly, production activities typically stretch over multiple periods and output in one period is often affect by the level produced in the previous period. As these examples indicate it is reasonable to argue that time series economic data may exhibit some kind of serial correlation.

The correlation between observations of a time series variable is captured by the autocorrelation function (ACF) which is given by:

$$ACF(s) = \frac{Cov(y_t, y_{t-s})}{\sigma_{y_t} \times \sigma_{y_{t-s}}}$$

Note that here t indexes the current period and s is an integer. For example, if $s = 1$, we are looking at correlation between y_t and y_{t-1}. This is known as **first order** serial correlation. Similarly, for $s = 2$, we get the **second order** serial correlation between y_t and y_{t-2}. In this ACF is a function of s and will give us a series of values of correlation of the current period with s past periods.

One of the main assumptions that affect the efficiency of the OLS estimator is the assumption of no serial correlation which forces the error term in our regression model to be independent across observations over time. Consider the following simple regression model:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

The classical assumption of no serial correlation implies that $Cor(\epsilon_t, \epsilon_{t-s}) = 0$, where t indexes current time period and s is an integer.

However, it is quite possible that the regression errors are actually correlated over time. For simplicity, lets assume that the error term in our model has first order serial correlation of the following form:

$$\epsilon_t = \rho \epsilon_{t-1} + u_t$$

Here ρ captures the first order serial correlation and is the slope parameter. u_t is a classical error term that satisfies all classical assumptions and hence is serially uncorrelated by definition. Now, depending on the sign of ρ the serial correlation can be positive or negative. In economics it is most common to observe positive serial correlation where the persistence in data ensures that a positive (negative) value of the regression error in one period is likely to be followed by another positive (negative) value. In contrast for negative serial correlation a positive error in one period is more likely to be followed by a positive error next period, and vice-versa. Figure 4.3 below shows the pattern of residuals for the two cases of positive and negative serial correlation in the regression error terms.

In general we can have higher order serial correlation in data and whether or not there is such correlation in our data is an empirical question.

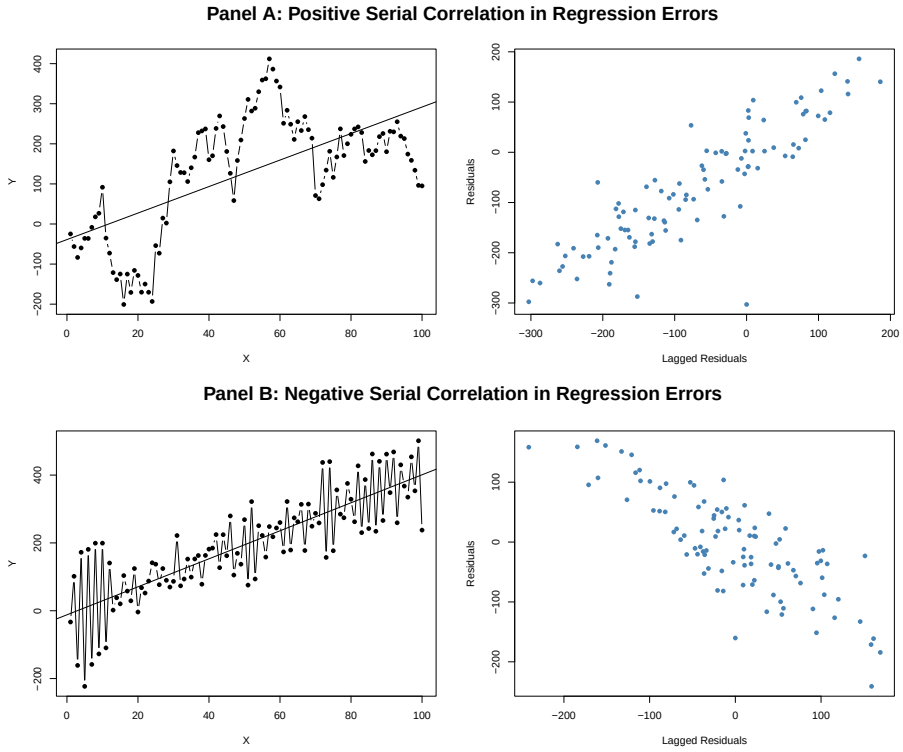


Figure 4.3: Serially correlated Errors

4.5.1 Consequences of Serial Correlation for the OLS estimator

Presence of serial correlation is a violation of the classical assumption and accordingly will affect the properties of the OLS estimator. In general, if there is serial correlation:

1. The OLS estimator of each β coefficient is still unbiased.
2. However, the OLS estimator is no longer efficient and the sample estimators of the standard errors of each β is incorrect. Consequently, we cannot conduct hypothesis testing on regression coefficients using the OLS estimator.

4.6 Testing for Serial Correlation in data

There are two tests for serial correlation that we will cover. The first one is only for testing the first order serial correlation. The second test is more general and can be used to test both the first order and higher order serial correlation.

4.6.1 Durbin-Watson test for first order serial correlation

One of the first tests for serial correlation is Durbin-Watson test. Note that this test can only be used to detect first order serial correlation. Consider the following regression model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \epsilon_t$$

Suppose we want to test whether the regression error term can be modeled as the first order serial correlation given by:

$$\epsilon_t = \rho \epsilon_{t-1} + u_t$$

Then, the test of serial correlation is basically testing whether ρ is 0 or not. There are two possibilities:

1. ρ is either 0 or positive:

$$H_0 : \rho = 0$$

$$H_A : \rho > 0$$

2. ρ is either 0 or negative:

$$H_0 : \rho = 0$$
$$H_A : \rho < 0$$

The Durbin-Watson test statistic, denoted by d, is given by:

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

where e_t denotes the residual from our regression model and T stands for sample size. Let $\hat{\rho}$ denotes the sample correlation coefficient between current and lagged residuals (i.e, e_t and e_{t-1}). Then we can show that:

$$d \approx 2(1 - \hat{\rho})$$

As a result there are three extreme values that this test statistic can take depending on $\hat{\rho}$:

$$d = \begin{cases} 0 & if\ \hat{\rho} = 1 \text{ (perfect positive serial correlation)} \\ 2 & if\ \hat{\rho} = 0 \text{ (no serial correlation)} \\ 4 & if\ \hat{\rho} = -1 \text{ (perfect negative serial correlation)} \end{cases}$$

Consequently, d is bounded between 0 and 4, and for positive serial correlation we expect d to take values sufficiently close to 0 whereas for negative serial correlation we expect d to take values sufficiently close to 4. Finally, a value of d close enough to 2 is indicative of no serial correlation.

In order to implement this test, we compare the estimated test statistic with 2 critical values: the lower limit denoted by dL and the upper limit denoted by dU. These values can be obtained from the Durbin-Watson distribution table using the information on the sample size and the number of independent variables in our regression model. The decision rules for the positive serial correlation as well as negative serial correlation tests are presented in Figure 4.4 and Figure 4.5.

4.6.2 Breusch-Godfrey (BG) test for serial correlation

In order to test for higher order serial correlations we can use the BG test which uses the OLS residuals to test for evidence of serial correlation. The procedure for this test is described below:

Step 1. Estimate the regression model using OLS and obtain residuals: e_t

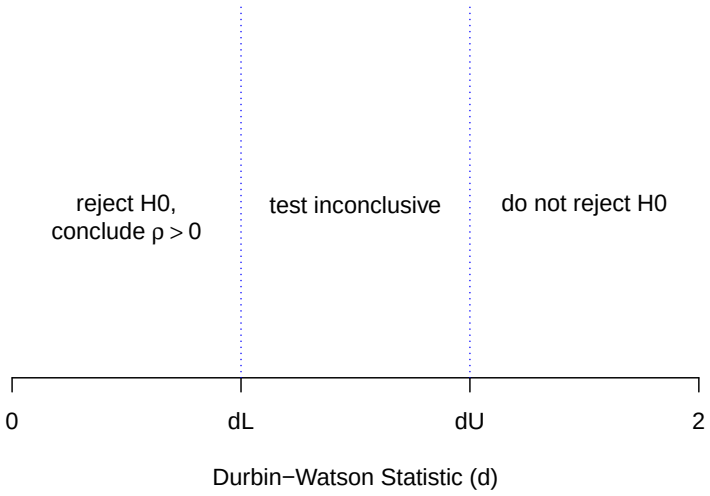


Figure 4.4: Decision-rule for Durbin-Watson Test of Positive Serial Correlation

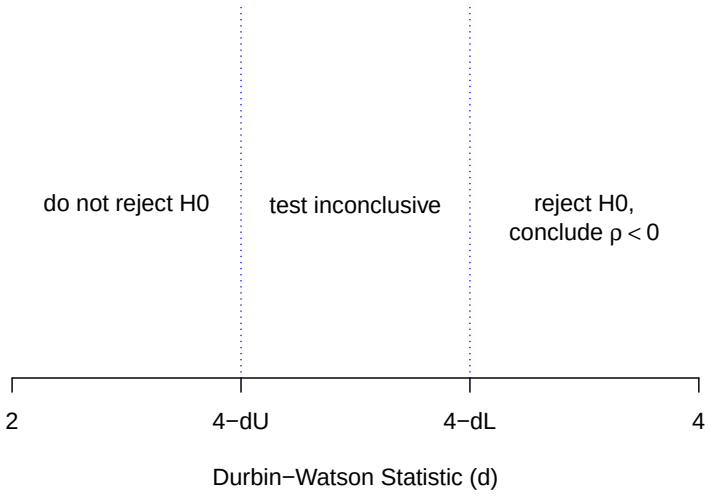


Figure 4.5: Decision-rule for Durbin-Watson Test of Negative Serial Correlation

Step 2: Estimate the BG regression model where the dependent variable is the residuals and independent variables are lagged values of this residual. The number of lagged residuals capture the order of serial correlation. In general for testing serial correlation up to order of p, we will estimate the following regression:

$$e_t = \alpha_0 + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + ... + \alpha_p e_{t-p} + u_t$$

Denote R^2_{BG} as the R-squared of this regression model.

Step 3: The serial correlation test is given by:

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = ... = \alpha_p = 0$$

$$H_A : Not\ H_0$$

The LM test statistic is given by:

$$LM = N \times R^2_{BG}$$

Under the null hypothesis, this test statistic follows Chi-square distribution with p degrees of freedom. If LM statistic is bigger than the critical value, we reject the null and conclude that there is serial correlation.

Example 4.2 (Testing for serial correlation)**.** Let us use the same example as the one we used for testing hetroscedasticity. We estimated a three factor model for the stock return of Apple using use monthly data from Jan-2007 through June 2019 (see Table 4.3.2 for OLS estimation results of this model).

We can implement both tests for serial correlation in R using the *lmtest* package. Table 4.6.2 below presents the results test-statistic for both serial correlation tests. For comparison, I am testing for first order serial correlation. the BG test clearly indicates presence of serial correlation as indicated by a p-value which is less than 0.05. For Durbin-Watson test, sample size is 150 and K=3. Using the Durbin-Watson table we get dL=1.584 and dU=1.665. Because d=0.75 is less than dL, we reject the null hypothesis of no serial correlation and conclude there is evidence for positive first order serial correlation.

	Durbin-Watson test	BG test
Test statistic	0.759	53.623
p-value	0.000	0.000

Two tests of Serial Correlation

4.7 Serial correlation robust standard errors

In case we find serial correlation in our data using OLS would lead to unbiased but inefficient estimators. The main consequence of serial correlation, as discussed above, is that the OLS standard errors for each β coefficient in our regression model are incorrect. In applied economic analysis it is common to replace the incorrect OLS standard errors with the **serial-correlation robust standard errors**. Table **??** reports the serial correlation robust standard errors for the regression model presented in Table 4.3.2. You can see minor differences in the standard errors for each β coefficient after correcting for serial correlation suggested by Newey-West (1994).

	Newey-West Robust Standard Errors
(Intercept)	0.017
Market Risk Premium	0.002
Size Premium	0.008
Volume Premium	0.006

Problems

Solutions

Chapter 5

Instrumental Variable Estimation

5.1 Endogeneity Problem

Suppose we have the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

In the above model, a crucial assumption for obtaining a causal effect of X on Y is that of **no endogeneity**, that is, the X variable is uncorrelated with the regression error (ϵ) and in that sense is **exogenous**. Figure 5.1. below shows the causal pathway between X and Y when X is exogenous.

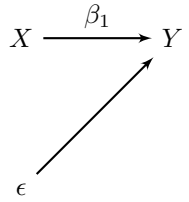


Figure 5.1: Exogenous X

However, in many real world applications that use a non-randomized research design this assumption is too strong and unlikely to be satisfied. here are three main sources of endogeneity (or $Cor(X_i, \epsilon_i) \neq 0$):

1. Omitted variable problem: a relevant X- variable is excluded from the regression model

2. Simultaneity problem or Reverse causality: theoretically it maybe possible to argue that X causes Y and Y causes X!

3. Measurement error in X variable: the data on X variable included in the model is noisy and we do not have a perfect measure of X

In each of the above cases, $Cor(X_i, \epsilon_i) \neq 0$ and hence X_i will be **endogenous**. Figure 5.2 below provides a graphical description of the breakdown of the causal link between X and Y when X is endogenous.

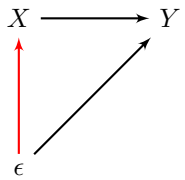


Figure 5.2: Endogenous X

As a consequence we no longer obtain an unbiased estimator of β_1 using OLS:

$$E(\widehat{\beta_1}) \neq \beta_1$$

In otherwords, when there is an endogenous regressor in the our model, using OLS would result in either an overestimation ($E(\widehat{\beta_1}) > \beta_1$) or an underestimation ($E(\widehat{\beta_1}) > \beta_1$).

5.2 Omitted variable bias

Although there are many reasons why we may face an endoegneous regressor, the most common source is an omitted variable where we exclude a relevant variable from our regression model. An omitted variable is a variable that should have been included in our regression model but is excluded either because of our ignorance or more likely due to a lack of measurement for such a variable. For example, consider the following simple model of wages:

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

where Y_i is the log of wages, X_{1i} is years of education and X_{2i} denotes a person's innate ability. However, it is very difficult to measure person's innate ability. Consequently, in practice our estimated regression model for wages excludes X_3 and is given by:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i^*$$

Here, $\epsilon_i^* = \beta_2 X_{2i} + \epsilon_i$. There are two conditions for X_{2i} to be an **omitted variable**:

1. It is a relevant variable, i.e, there is a non-zero effect of X_3 on Y (implied by $\beta_3 \neq 0$)
2. It is correlated with at least one of the included X variables in the model. This will lead to correlation between the included X variables and the error term with the omitted variable.

In our example, it is reasonable to argue that a person with greater innate ability will be more likely to get more education, i.e,$Cor(X_{1i}, X_{2i}) > 0$. Further, a person's innate ability should also positively affect her wage, i.e., $\beta_2 > 0$. As a result, both of the above conditions are satisfied implying innate ability indeed is an omitted variable in our estimated model.

5.2.1 Consequence of omitted variable on OLS estimator

The main consequence of an omitted variable is that the included X variables are no longer exogenous and are correlated with the error term. In our example this means that $Cor(X_{1i}, \epsilon_i^*) \neq 0$. As a result the OLS estimator of education is no longer unbiased:

$$E(\widehat{\beta_1}) \neq \beta_1$$

Although, we cannot estimate the magnitude of the bias but in some cases it may be possible to guess the sign of the bias. In our example, the sign of this bias will depend on the signs of β_2 and $Cor(X_{1i}, X_{2i})$. Specifically, in the case where only 1 variable is omitted and only 1 variable is included in the model (like our example) we get the following 4 possibilities:

	$Cor(X_1, X_2) > 0$	$Cor(X_1, X_2) < 0$
$\beta_2 > 0$	$E(\widehat{\beta_1}) > \beta_1$ (positive bias)	$E(\widehat{\beta_1}) < \beta_1$ (positive bias)
$\beta_2 < 0$	$E(\widehat{\beta_1}) < \beta_1$ (positive bias)	$E(\widehat{\beta_1}) > \beta_1$ (negative bias)

In our case, higher innate ability should increase wages ($\beta_2 > 0$) and people with greater innate ability also have higher levels of education ($Cor(X_1, X_2) > 0$). Hence, the OLS estimator of education's effect on wages will be positively biased–part of the effect is due to innate ability which is omitted from the model.

In order to resolve this problem we will need to use Intstrumental variable (IV) estimator which attempts to exploit exogenous variation in the endogenous independent variable. We will cover this method in advanced econometrics course.

5.3 IV estimation

One way to deal with endogeneity is to **instrument** for the endogenous X-variable: **Instrumental Variable (IV)** estimation. An instrumental variable (denoted by Z) has the property that that changes in Z are associated with changes in X but only affect Y indirectly through X. Figure 5.3 below describes how using an instrument identifies causal effect of an endogenous regressor (X) on Y.

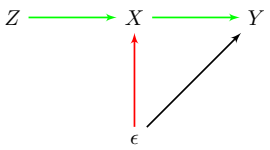


Figure 5.3: Using Z as an instrument for endogenous X

An instrument Z is a valid instrument if it meets the following two conditions:

```
1. It is **relevant** in the sense that it is related to endogenous X variable, i.e., $

2. It is **exogenous** in the sense that is uncorrelated with the error term, i.e., $C
```

5.4 Case I: One endogenous regressor and one instrument

In this section we consider a simple case of only one endogenous regressor that has exactly one available instrument. Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i$$

Suppose we know that X is endogenous, i.e., $Cor(X_i, \epsilon_i) \neq 0$. Further assume that we are able to find a valid instrument denoted by Z_i. In this case our regression model is **just-identified** because the number of instruments equals number of endogenous variables. One way to derive an instrumental variable (IV) estimator is two-stage least squares (2SLS). This procedure isolates the exogenous variation in X which is then used to estimate β_1. There are two stages to 2SLS:

1. First-stage regression: Here we regress X on Z and obtain the predicted value of X using OLS. This first-stage regression is given by:

$$X_i = \pi_0 + \pi_1 \cdot Z_i + u_i$$

The predicted value from this regression is given by:

$$\widehat{X_i} = \widehat{\pi_o} + \widehat{\pi_1} \cdot Z_i$$

Note that by construction, $\widehat{X_i}$ is that part of X_i that is exogenous, assuming our instrument is valid.

2. Second-stage regression: Here we replace X_i in our original regression with $\widehat{X_i}$ and estimate the model with OLS:

$$Y_i = \beta_0 + \beta_1 \cdot \widehat{X_i} + \epsilon_i$$

The above regression yields $\hat{\beta}_1^{2SLS}$ or $\hat{\beta}_1^{IV}$ which will be unbiased if our instrument is both relevant and exogenous.

5.4.1 Example 1: Return to Schooling

Suppose we are interesting in estimating the causal effect of years of education (X) on wages (Y). As discussed earlier, due to missing information on innate ability, we have an endgoeneity problem. We need to find an instrument Z that is correlated with schooling, uncorrelated with ability. One such candidate used in the literature is distance from school or college. The idea being proximity to an educational institution can positively affect schooling (relevance of Z) but in theory should not correlate with a person's innate ability (exogeneity of Z).

5.4.2 Example 2: Estimating Demand for Butter

One of the first applications of IV estimation was to the elasticity of demand for butter (and other agricultural products). Consider the following regression model:

$$ln(Q_i) = \beta_0 + \beta_1 \cdot ln(P_i) + \epsilon_i$$

Here, Q denotes quantity demanded of butter and P denotes price of butter. β_1 is the percent change in quantity demanded caused by 1% change in its price, i.e., elasticity of demand for butter. One issue with estimating the above model using OLS is that of reverse causality. Economic theory suggests that any pair

of price and quantity observed in the market is an equlibrium point where demand intersects supply. As a result, both quantity and price are simultaneously determined leading to reverse causality where price causes quantity and quantity causes price. To break this endogeneity we need to find an instrument for endogenous price. One such instrument is a supply shifter that only affects the supply curve. e.g. rainfall in regions that produce dairy products. It is plausibly relevant (affects price by reducing quantity of butter supplied due to insufficient rain) and exogenous (should not directly affect demand). Formally,

Stage 1: Regress price on rainfall

$$ln(Pi) = \pi_0 + \pi_1 \cdot Z_i + \epsilon_i$$

Stage 2: Replace price with predicted price from the first-stage regression:

$$ln(Qi) = \beta_0 + \beta_1 \cdot \widehat{ln(P_i)} + \epsilon_i$$

The estimated slope coefficient from the second-stage regression is our IV estimator of elasticitty of demand for butter that is unbiased/causal.

5.5 Case II: One endogenous regressor with many available instruments

Often times we are faced with a situation where we have many candidate instruments for an endogenous regressor. Lets go back to our example of demand for butter. As we discussed earlier, one candidate instrument is rainfall as it affects supply which in turn affects oprice. But there are many other variables that can affect price through their effect on supply. For example, use of high-yield cows can affect supply of milk which in turns affects supply of butter and hence its price. In this case we have two instruments Z_1 (rainfall) and Z_2 (high-yield cows) for the endogenous price. In otherwords our model is **overidentified**. How does this impact our IV estimator?

Stage 1: The first-stage regression now utilizes both instruments:

$$ln(Pi) = \pi_0 + \pi_1 \cdot Z_{1i} + \pi_2 \cdot Z_{2i} + \epsilon_i$$

Stage 2: The second-stage regression stays the same:

$$ln(Qi) = \beta_0 + \beta_1 \cdot \widehat{ln(P_i)} + \epsilon_i$$

5.6 IV estimation in a multiple regression framework

Our discussion so far has focused on a simple regression with only one X variable that could be endogenous. However, in many applications we may have other X-variables in the model. There are two interesting scenarios to consider in the multiple regression setting:

1. Only one endogenous regressor so that all other X variables included in the model are exogenous. Here we need at least one instrument to estimate the causal effect of the endogenous regressor.
2. Multiple endogenous regressors in the model. Here we need at least as many instruments as the number of endogenous regressors in the model.

Let us reconsider the return to schooling example. Consider the following multiple regression model:

$$ln(wages_i) = \beta_0 + \beta_1 \cdot educ_i + \beta_2 ln(exper_i) + \epsilon_i$$

Here it is reasonable to argue that the omitted variable (innate ability) affects education but may have negligible impact on years of experience. In this case, education is the endogenous regressor that needs an instrument whereas experience is exogenous and hence does not require an instrument. Let Z denotes proximity to educational institution, an instrument for years of education. Then the 2SLS is given by:

Stage 1: In the first stage we regresss education on the instrument and experience:

$$educ_i = \pi_0 + \pi_1 \cdot Z_i + \gamma_1 \cdot ln(exper_i) + u_i$$

Stage 2: In the second stage we replace education with the predicted value from the first stage:

$$ln(wages_i) = \beta_0 + \beta_1 \cdot \widehat{educ_i} + \beta_2 ln(exper_i) + \epsilon_i$$

```
Key thing to note here is that the first stage regression must include all the instruments and al
```

The most general case includes multiple endogenous regressors and many exogneous regressors. This case will not be covered in this chapter but is quite easy to follow from the previous discussion.

5.7 Strength and Exogneity of the Instrument

How do we know whether the instrument we have chosen is a good instrument? There are two related issues here:

1. Weak instrument problem: happens when our instrument(s) are only weakly associated with the endogenous regressor. This implies our instrument does not satisfy the relevancy condition.
2. Endogenous instrument problem: occurs when our instrument(s) are not exogenous and in that sense fail to isolate exogenous variation in the endogenous X variable.

5.7.1 Weak Instruments

If our chosen set of instruments are weak, then the first-stage regression can be used to statistically test for such a scenario. Consider the following example:

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

Suppose we know only X_1 is endogenous and we propose three possible instruments Z_1, Z_2, and Z_3 for it. The first-stage regression is given by:

$$X_{1i} = \pi_0 + \pi_1 \cdot Z_{1i} + \pi_2 \cdot Z_{2i} + \pi_3 \cdot Z_{3i} + \gamma_1 \cdot X_{2i} + u_i$$

Then, the test of weak instruments is given by:

$$H_0 : \pi_1 = \pi_2 = \pi_3 = 0 \rightarrow \text{instruments weak}$$

$$H_A : \text{Not } H_0$$

The F-statistic for the above hypothesis is called the **first-stage F stat.** and larger value is desirable as it would increase the likelihood of rejecting the null hypothesis of weak instrunents. It is common to report the value of this test statistic when presenting the results of the IV estimation. For the single endogenous regressor case Stock and Yogo (2005) show that values of the first-stage stat. greater than 10 can serve as a reliable rule-of-thumb in assessing strength of the instruments.

5.7.2 Exogeneity of Instruments

When we have more than one instrument for the same endogneous regressor, we can exploit the overid.

Problems

Solutions

Chapter 6

Discrete Choice Model

6.1 Binary Dependent Variable

So far we have discussed how we can use a dummy variable as a regressor in our model. For example, if we want to capture the effect of sex on wages, we can simply include a binary variable capturing sex of a worker in our regression model. However, some times we may want to use a dummy variable as our dependent variable. Consider the following scenario. Suppose you are managing financial risk for a major bank and you have been tasked to explain what factors leads to a default on a personal loan. In this case, the dependent variable of interest is categorical in nature. A person who borrowed can either default or not. In this case our outcome of interest is a binary variable:

$$Y_i = \begin{cases} 1 & \text{if observation } i \text{ defaults on loan} \\ 0 & \text{otherwise} \end{cases}$$

Suppose one of the important factor affecting default is the financial wellbeing of a person, denoted by X. Then, our simple regression model is given by:

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i$$

In this chapter we will learn how do we estimate this model and more importantly, why using OLS may not be a good idea when our dependent variable is a dummy variable.

Problems

Solutions

Appendix A

Review of Differential Calculus and Optimization

Given that all students must have taken a course in calculus before enrolling for this class, it is assumed that everyone in the class is comfortable with concepts such as derivatives, partial derivatives, and optimization. In this chapter, I will provide a brief review of some concepts that are most pertinent for Econometrics. I strongly encourage that you read your lecture notes for Calculus if you find it difficult to follow the material presented in this chapter.

A.1 Derivative of a single variable function

Definition A.1 (Derivative of a function)**.** Consider the following function, $y = f(x)$. The *derivative* of this function measures the rate of change in y caused by a change in x.

There are two alternative notations for the derivative of y with respect to x: $f'(x)$ or $\frac{dy}{dx}$.

The derivative of a function is very closely related to the concept of *slope* of a function. Let Δ denotes change in a variable. Then, by definition, the slope of y with respect to x is given by:

$$slope = \frac{\Delta y}{\Delta x}$$

The derivative of y with respect to x is the limit value of the slope as $\Delta x \to 0$. Hence,

$$\frac{dy}{dx} \text{ or } f'(x) = \lim_{\Delta x \to 0} \left(\frac{\Delta y}{\Delta x} \right)$$

A.1.1 Rules of Differentiation

1. Derivative of a constant is 0.
2. Derivative of a function multiplied by a constant is constant times the derivative of the function:

$$\frac{d}{dx}[\ a \times f(x)] = a \times f'(x)$$

where it is assumed that a is an constant.

3. Addition rule:

$$\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$$

4. Subtraction rule:

$$\frac{d}{dx}[f(x) - g(x)] = f'(x) - g'(x)$$

5. Product rule:

$$\frac{d}{dx}[f(x) \times g(x)] = f(x) \times g'(x) + g(x) \times f'(x)$$

6. Quotient rule:

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x) \times g(x) - g'(x) \times f(x)}{g(x)^2}$$

7. Chain rule:

$$\frac{d}{dx}[f(g(x))] = f'(g(x)) \times g'(x)$$

8. Derivative of some common functions:

a. Power function: $f(x) = x^a$. Then,

$$f'(x) = a \times x^{a-1}$$

b. Natural log function: $f(x) = ln(x)$. Then,

$$f'(x) = \frac{1}{x}$$

c. Exponential function: $f(x) = e^x$

$$f'(x) = e^x$$

A.2 Second derivative and non-linearity

Definition A.2 (Second derivative of a function)**.** Consider the following function, $y = f(x)$. The *second derivative* of this function measures the change in the rate of change of this function. Formally it is denoted by $f''(x)$ or $\frac{d^2y}{dx^2}$.

The second derivative measures the *curvature* of the function and hence can be used to distinguish a *linear* function from a *non-linear* function. By definition, a linear function has a constant slope implying the its second derivative must be zero.

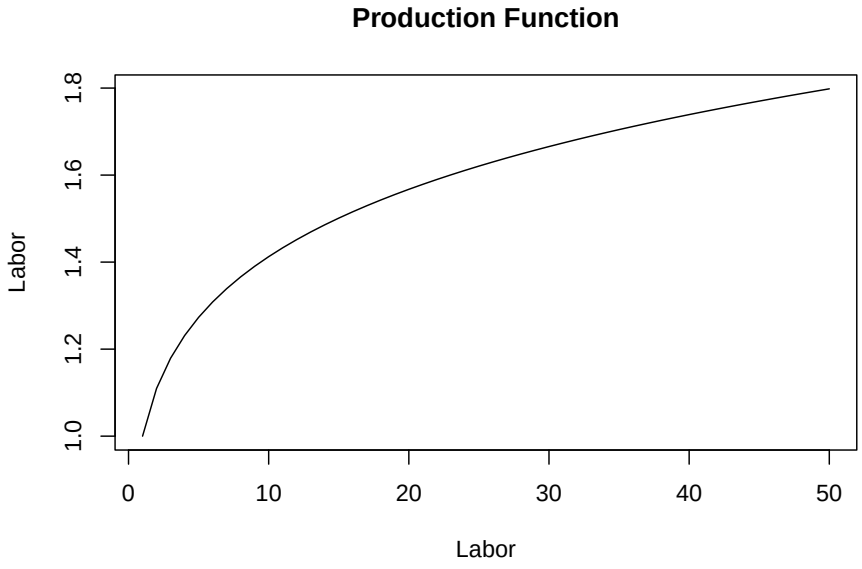
Example A.1. For example, consider the following linear function:

$$f(x) = mx + b$$

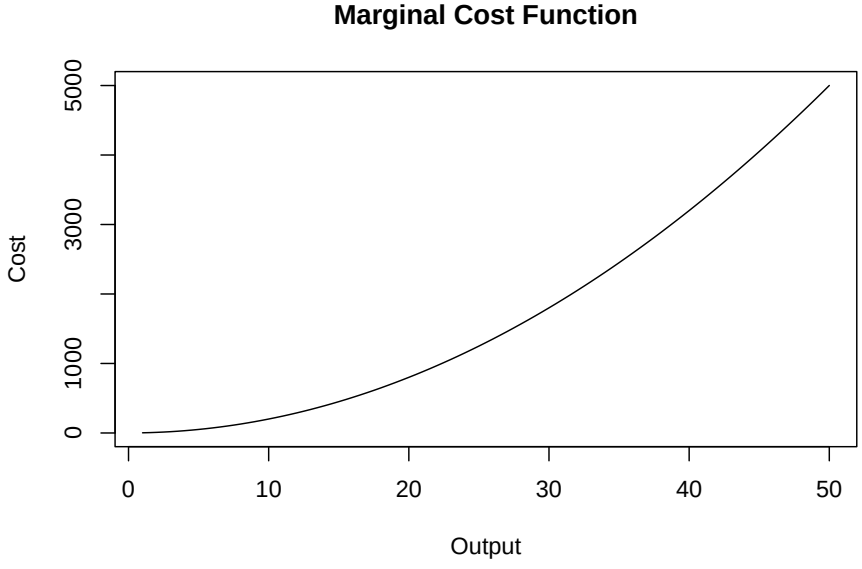
Here$f'(x) = m$ and $f''(x) = 0$.

A non-linear function will have a non-zero second derivative. There are only two possibilities:

1. $f''(x) < 0$. In this case we have a concave relationship. An example from economics is the production function where the relationship between output and input is concave.



2. $f''(x) >$. In this case we have a convex relationship. An example from economics is the marginal cost function where the relationship between cost of production and level of output can be convex.



A.3 Partial derivatives: Multi-variable functions

Ceteris paribus aka *holding other things equal* is one of the key concepts used in Economic analysis. A *partial derivative* is a mathematical counterpart of this assumption.

Definition A.3 (Partial Derivative)**.** Consider a function of n-variables given by $y = f(x_1, x_2, x_3, ..., x_n)$. Then, there are n partial derivatives of this function that can be obtained by taking derivative with respect to one of the x-variables, holding all other constant. Formally, the partial derivative of y with respect to x_i is denoted by f_{x_i} or $\frac{\partial y}{\partial x_i}$.

Example A.2. Consider the following 3-variable function:

$$y = ln(x_1) + x_1 \times x_2 + 3x_2^2 + x_1 \times x_3 + ln(x_3)$$

Then we can compute three partial derivatives of this function:

- Partial derivative of y with respect to x_1, treating x_2 and x_3 as constants:

$$\frac{\partial y}{\partial x_1} = \frac{1}{x_1} + x_2 + x_3$$

- Partial derivative of y with respect to x_2, treating x_1 and x_3 as constants:

$$\frac{\partial y}{\partial x_2} = x_1 + 6x_2$$

- Partial derivative of y with respect to x_3, treating x_1 and x_2 as constants:

$$\frac{\partial y}{\partial x_3} = x_1 + \frac{1}{x_3}$$

Example A.3 (Cobb-Douglas Production Function)**.** One of the most used functional form for the production function is the Cobb-Douglas production function. Suppose you have two inputs: labor (L) and capital (K). Let Y denotes output. Then, the Cobb-Douglas production function is given by:

$$Y = L^{\beta_1} K^{\beta_2}$$

Now, output can change because we change our labor input or our capital input. In each case, we are thinking about a change in output caused by change in one input, holding the other input constant. This is exactly what a partial derivative captures! In what follows next we will use two mathematical concepts to further our understanding of economics of production:

1. Change in natural logs of a variable approximates percent change in that variable. Formally, $\Delta ln(x) \times 100 \approx$ % change in x. Hence, it is often useful to express economic relationships in natural logs. The Cobb-Douglas production function in natural logs is given by:

$$ln(Y) = \beta_1 \times ln(L) + \beta_2 \times ln(K)$$

2. The partial derivative of the above equation gives us **elasticity of output** with respect to each input.

   a. Output elasticity of Labor:

$$\frac{\% \text{ change in Y}}{\% \text{ change in L}} = \frac{\partial ln(Y) \times 100}{\partial ln(L) \times 100} = \beta_1$$

   b. Output elasticity of Capital:

$$\frac{\% \text{ change in Y}}{\% \text{ change in K}} = \frac{\partial ln(Y) \times 100}{\partial ln(K) \times 100} = \beta_2$$

Note that we can also infer whether production is subject to increasing, decreasing, or constant returns to scale from the numerical values assigned to β_1 and β_2. Returns to scale is simply the sum of output elasiticities with respect to labor and capital:

$$\text{Returns to scale} = \frac{\% \text{ change in Y}}{\% \text{ change in L}} + \frac{\% \text{ change in Y}}{\% \text{ change in K}} = \beta_1 + \beta_2$$

Hence, we obtain constant returns to scale as long as $\beta_1 + \beta_2 = 1$. We get decreasing returns to scale if $\beta_1 + \beta_2 < 1$. Finally, increasing returns to scale require $\beta_1 + \beta_2 > 1$.

A.4 Optimization

In Economics it is often assumed that rational individuals *optimize*. For instance, firms seek to maximize profits (or minimize costs) and households seek to maximize utility. Mathematically, this is equivalent to finding **extreme** values of an **objective function**.

Example A.4. Consider a firm that is choosing a level of output (q) to maximize its profits. By definition, profits are total revenue $R(q)$ minus total cost $C(q)$. The resulting profit function $\pi(q)$ is the firm's objective function and q is the control variable:

$$\pi(q) = R(q) - C(q)$$

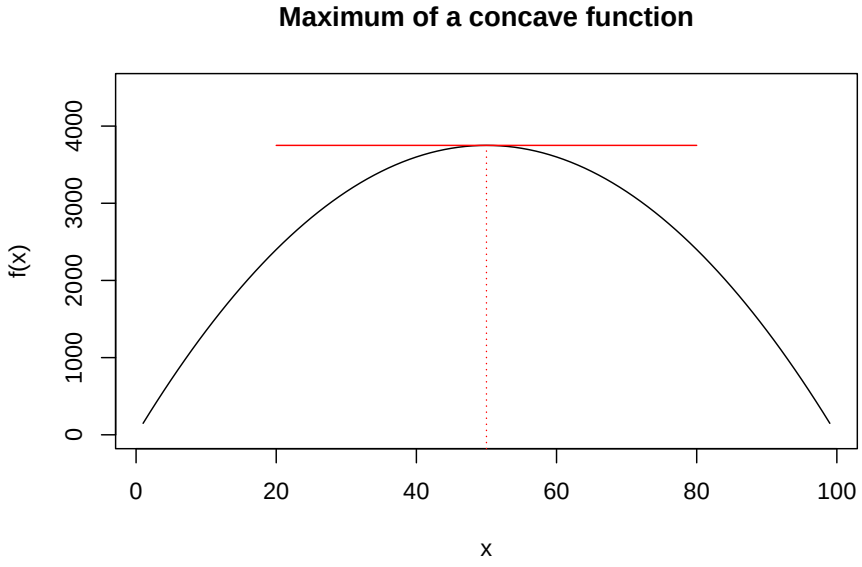
The firm will choose a value of q that will maximize its profits. Mathematically, this can be written as:

$$\max_q \pi(q)$$

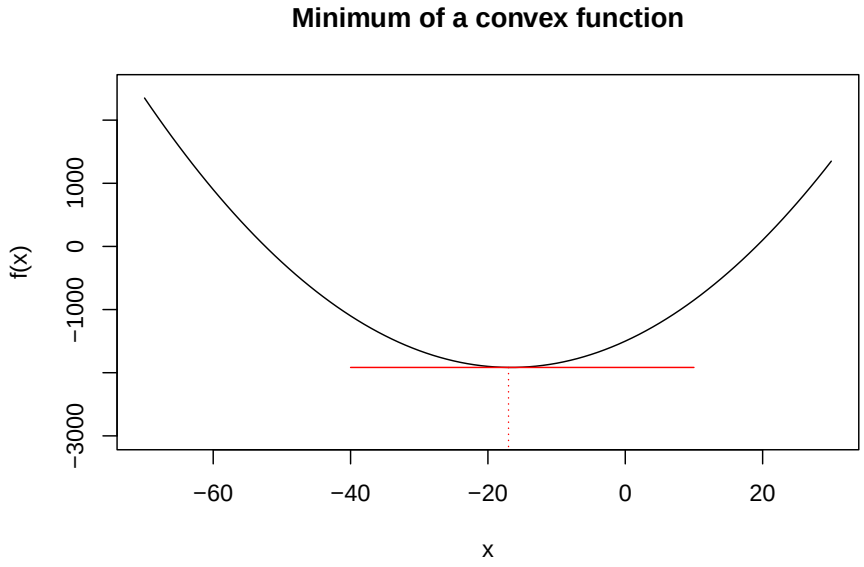
One way to solve this problem, is to assume a functional form for profits and evaluate this function for all possible values of q. Then, select the value of q that yields highest value for profits. This approach is called **numerical optimization** and is often used for complicated objective functions. But in many cases, we can use calculus and obtain an *analytical* solution for the optimization problem.

Formally, suppose the objective function is denoted by $f(x)$ and assume that this function is continuous and twice differentiable. Then,

1. x^* is a maximizer if $f(x^*) \geq f(x)$ for all $x \neq x^*$. Note that at this point the slope of the tangent to the function is 0, i.e., $f'(x^*) = 0$. This is the **first order condition (foc)** for obtaining a maximum. The graph below illustrates the maximum of a generic function. Note that the slope of the function changes sign from positive to negative around x^*. This will give us the **second order condition** for obtaining a maximum.



2. x^* is a minimizer if $f(x^*) \leq f(x)$ for all $x \neq x^*$. Note that at this point the slope of the tangent to the function is 0, i.e., $f'(x^*) = 0$. This is the **first order condition (foc)** for obtaining a minimum. The graph below illustrates the minimum of a generic function. Note that the slope of the function changes sign from negative to positive around x^*. This will give us the **second order condition** for obtaining a minimum.



Note for a maximum, Similarly, we can now outline the steps for computing a maximum or minimum of a given function.

1. First-order condition: Compute the first derivative of the function and equate it to 0. The solution to this equation gives us x^*:

$$f'(x^*) = 0$$

2. Second-order condition: Compute the second derivative of the function and evaluate it at x^*.
   a. If $f''(x^*) < 0$, then x^* is a maxmizer.
   b. If $f''(x^*) > 0$, then x^* is a minimizer.

Example A.5 (Single variable optimization example)**.** Consider a firm that produces a single good q and sells it at a price of \$10 per unit. The cost of production is given by:

$$C(q) = 2q + 5 + 0.1q^2$$

At what level of output would profits be maximized?

Solution. The profit of a firm is revenue minus cost:

$$\pi(q) = R(q) - C(q) = 10q - 2q - 5 - 0.1q^2 = 8q - 5 - 0.1q^2$$

Hence, we want to solve the following problem:

$$\max_q \pi(q)$$

The first order condition is given by:

$$\pi'(q) = 0 \Rightarrow 8 - 0.2q = 0 \rightarrow q^* = 40$$

The second order condition is given by:

$$\pi''(q) = -0.2 < 0$$

Hence, $q^* = 10$ maximizes the profits. The maximum level of profits is given by $\pi(q^*) = 8 \times 40 - 5 - 0.1 \times 40^2 = 155$.

Note that the above process can be easily applied to multivariate functions. In that case there will be one first order condition for every control variable.

Example A.6 (Multi-variable optimization example)**.** Consider a two-variable function:

$$f(x_1, x_2) = 2x_1x_2 + \frac{100}{x_1} - 4x_2^2$$

Solve the following minimization problem:

$$\min_{x_1, x_2} f(x_1, x_2)$$

Solution. Now we have two first order conditions:

$$f_{x_1}(x_1, x_2) = 0 \Rightarrow 2x_2 - \frac{100}{x_1^2} = 0$$

$$f_{x_2}(x_1, x_2) = 0 \Rightarrow 2x_1 - 8x_2 = 0$$

So we have two equations in two unknowns. You can show that $x_1* = 5.84$ and $x_2^* = 1.46$. The minimum of this function is given by $f(x_1^*, x_2^*) = 2 \times 5.84 * 1.46 + \frac{100}{5.84} - 4 \times 1.46^2 = 25.65$.

Problems

Exercise A.1. Compute the derivative of the following functions.

a. $f(x) = 2x^2$

b. $f(x) = 2x^2 + ln(x)$

c. $f(x) = e^{ax}$

d. $f(x) = (2x + x^2)^3$

e. $f(x) = ln(5x + x^2)$

f. $f(x) = \dfrac{x + ln(x)}{x^3}$

Exercise A.2. Compute the second derivative of each function given in Exercise 1.1.

Exercise A.3. Compute the partial derivative for each variable for the following functions:

a. $f(x_1, x_2, x_3) = 4x_1^3 x_2 - e^{x_3} x_1 + 3x_2$

b. $f(x_1, x_2) = \dfrac{2x_1 + 3x_2}{4x_1^3 - 7x_1 x_2}$

c. $f(x, y) = ln(y^2) - ln(x) + 2ln\left(\dfrac{x}{y}\right)$

d. $f(x, y) = 2x^{0.4} y^{0.8} + 2x$

Exercise A.4. Solve the following optimization problems. In each case compute the maximizer(s) (or minimizer(s)) for the function as well as the optimum value of the function.

a. $\max\limits_{x} f(x) = 3ln(x) - 0.5x + 4$

b. $\min\limits_{x,y} f(x, y) = 2xy + \dfrac{2000}{x} + \dfrac{2000}{y}$

c. $\max\limits_{x} f(x) = ax^{0.5} - bx + 4$

Solutions

Exercise 1.1:

a. $f'(x) = 4x$

b. $f'(x) = 4x + \frac{1}{x}$

c. $f'(x) = ae^{ax}$

d. $f'(x) = 3(2x + x^2)^2(2 + 2x)$

e. $f'(x) = \dfrac{5 + 2x}{5x + x^2}$

f. $f'(x) = \dfrac{x^3(1 + 1/x) - 3x^2(x + ln(x))}{x^6} = \dfrac{x^2 - 2x^3 - 3x^2 ln(x)}{x^6}$

Exercise 1.2:

a. $f''(x) = 4$

b. $f''(x) = 4 - x^{-2}$

c. $f''(x) = a^2e^{ax}$

d. $f''(x) = 6(2x + x^2)^2 + 6(2x + x^2)(2 + 2x)^2$

e. $f''(x) = \dfrac{2(5x + x^2) - (5 + 2x)^2}{(5x + x^2)^2}$

f. $f''(x) = \dfrac{x^6(-x - 6x^2 - 6xln(x)) - 6x^5(x^2 - 2x^3 - 3x^2 ln(x))}{x^{12}}$

Exercise 1.3:

a.

$$\frac{\partial f}{\partial x_1} = 12x_1^2x_2 - e^{x_3}$$

$$\frac{\partial f}{\partial x_2} = 4x_1^3 + 3$$

$$\frac{\partial f}{\partial x_3} = -e^{x_3}x_1$$

b.

$$\frac{\partial f}{\partial x_1} = \frac{2(4x_1^3 - 7x_1x_2) - (2x_1 + 3_x2)(12x_1^2 - 7x_2)}{(4x_1^3 - 7x_1x_2)^2}$$

$$\frac{\partial f}{\partial x_2} = \frac{3(4x_1^3 - 7x_1x_2) + 7x_1(2x_1 + 3_x2)}{(4x_1^3 - 7x_1x_2)^2}$$

c. [Hint: simplify using properties of logs before taking the partial derivative.]

$$\frac{\partial f}{\partial x} = \frac{1}{x}$$

$$\frac{\partial f}{\partial y} = 0$$

d.

$$\frac{\partial f}{\partial x} = 0.8x^{-0.6}y^{0.8} + 2$$

$$\frac{\partial f}{\partial y} = 1.6x^{0.4}y^{-0.2}$$

Exercise 1.4:

a. First order condition for maximum gives us:

$$f'(x) = 0 \Rightarrow \frac{3}{x} - 0.5 = 0 \Rightarrow x^* = 6$$

The maximum of the function is given by:

$$f(x^*) = 3 \times ln(6) - 0.5 \times 6 + 4 = 6.375$$

b. Now we have the following two first order conditions:

$$\frac{\partial f}{\partial x} = 2y - \frac{2000}{x^2} = 0$$

$$\frac{\partial f}{\partial y} = 2x - \frac{2000}{y^2} = 0$$

Solving for x and y gives us $x^* = 10$ and $y^* = 10$. The minimum of the function is given by:

$$f(x^*, y^*) = 2 \times 10 \times 10 + \frac{2000}{10} + \frac{2000}{10} = 600$$

c. The first order condition gives us:

$$f'(x) = 0.5ax^{-0.5} - b = 0 \Rightarrow x^* = \left(\frac{a}{2b}\right)^2$$

The maximum of the function is given by:

$$f(x^*) = \frac{a^2}{2b} - \frac{a}{2} + 4$$

Appendix B

Review of Probability and Statistics

Given that all students must have taken a course in statistics before enrolling for this class, it is assumed that everyone in the class is comfortable with concepts such probability, expected value, measures of central tendency, hypothesis testing etc. In this chapter, I will provide a brief review of some concepts that are most pertinent for Econometrics. I strongly encourage that you read your lecture notes for Statistics if you find it difficult to follow the material presented in this chapter.

B.1 Probability

We begin with a brief review of probability theory. To define probability we first need to develop an understanding of what we mean by *experiment*, *sample space*, and *event* in statistics.

Definition B.1 (Experiment)**.** An experiment is a process with an uncertain observable outcome. e.g. Toss of a coin can have two possible outcomes, heads or tails.

Definition B.2 (Sample Space)**.** The sample space is the set of all possible outcomes of an experiment. I will denote it by S. If we toss a coin then $S = \{Heads, Tails\}$.

Definition B.3 (Event)**.** An event is a subset of the sample space. I will denote it by E. If we toss a coin and Heads shows up then $E = Heads$.

Now, we can define probability, which is a function that assigns a numerical value to the chance of an event occurring among all possible events in the sample space.

Definition B.4 (Probability)**.** A function P is called a probability function if:

1. For any given event, E, $0 \leq P(E) <\leq 1$.
2. Suppose there are N possible events in S, i.e., $S = \{E_1, E_2, E_3, .., E_N\}$. Then,
$$P(E_1) + P(E_2) + P(E_3) + ... + P(E_N) = 1$$
3. Consider an event E. Then,
$$P(\neg E) = 1 - P(E)$$
3. If we have two disjoint events A and B, then:
   a. $P(A \cup B) = P(A) + P(B)$
   b. $P(A \cap B) = 0$
4. If we have two non-disjoint events A and B, then:
   a. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   b. $P(A \cap B) = P(A) \times P(B|A)$
5. If we have two independent events A and B, then:
$$P(A \cap B) = P(A) \times P(B)$$
6. Bayes rule:
$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

where $P(B) = P(B|A) \times P(A) + P(B|\neg A) \times P(\neg A)$

One such probability function is:

$$P(E) = \frac{\text{Number of outcomes in E}}{\text{Number of outcomes in S}} \tag{B.1}$$

Example B.1. Consider a fair six-sided dice. The probability of obtaining an odd number if this dice is rolled once is given by 0.5. To see this, note that the event here is obtaining an odd number when a dice is rolled. Hence, $E = \{1, 3, 5\}$. Also, $S = \{1, 2, 3, 4, 5, 6\}$. Using this, we get:

$$P(E) = \frac{3}{6} = 0.5 \tag{B.2}$$

B.2 Random Variable

One of the most important applications of statistics is to resolve the randomness that is inherent in most economic choices. For example, the outcome of your college major is a random variable with many possible values. Most economic variables can be thought of as **random variables** that have many possible values which are unknown until they are realized. We will begin by formally defining a random variable.

Definition B.5 (Random Variable)**.** A random variable is a numerical representation of outcomes of an experiment. For example, in the example of a toss of a coin, suppose you win $10 if heads shows and you lose $5 if tails shows. In this case, tossing the coin was the experiment, and winnings from this game is the random variable with two possible values: $10 and -$5.

There are two types of random variables.

1. Discrete random variable: takes finite number of values. e.g. GPA points earned in Econ 385.
2. Continuous random variable: can take any value on the number line. e.g. GDP in the last quarter of 2019.

B.3 Probability distribution

By definition a random variable can take many *possible values*. In statistics a function that provides the probabilities of different realizations of a random variable is called its **probability distribution**.

B.3.1 Probability distribution of a discrete random variable

For a discrete random variable the probability distribution is simply the list of all possible values this variable can and their corresponding probabilities. Let X be a discrete random variable with n possible values give by $\{x_1, x_2, x_3, .., x_n\}$. Let p_i denotes that probability that $X = x_i$. Then, the probability distribution function of this random variable is given by:

X	p(X)
x_1	p_1
x_2	p_2
x_3	p_3

X	p(X)
⋮	⋮
x_n	p_n

Example B.2 (Grade Distribution)**.** A typical grade distribution is an example of a discrete random variable. Consider the following grade distribution:

GPA	Percent of Students
0	10%
1	20%
2	40%
3	20%
4	10%

Note that every GPA point corresponds to a letter grade. From the perspective of the student, X is the random variable that is his letter grade, and the above distribution gives the probability of obtaining a particular letter grade. We can plot this simple probability distribution as follows:

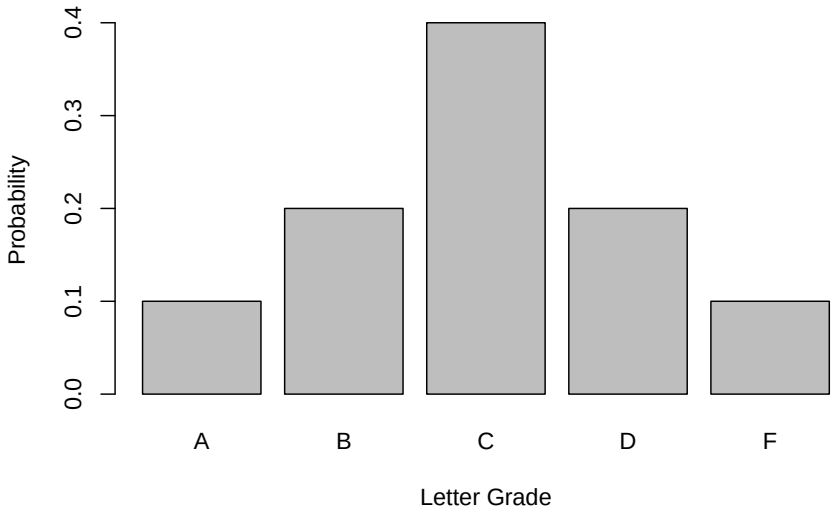


Figure B.1: Probability Distribution of Letter Grades

We can use the probability distribution of a discrete random variable in two different ways.

1. We can compute the probability of the random variable taking an exact value. This is known as the **probability mass function (p.m.f)** and is denoted by $f(x)$:

$$f(x) = P(X = x)$$

For example, the probability of obtaining a letter grade of C or $P(X = 2)$ is 0.4 or 40%.

2. We can also infer the probability that a discrete random variable will be less than or equal to a certain value by cumulatively adding the probabilities. Formally, we can compute the **cumulative probability distribution (c.d.f)** which is denoted by $F(x)$:

$$F(x) = P(X \leq X)$$

Going back to our grade distribution example, we can add the column of cumulative probabilities to obtain the $c.d.f$:

Grade	Percent of Students	$F(x)$
0	10%	10%
1	20%	30%
2	40%	70%
3	20%	90%
4	10%	100%

So for example, we can infer that the probability of obtaining the letter grade of C or lower i.e, $P(X \leq 2)$ is 0.7 or 70% which is obtained by adding the probabilities of obtaining letter grades of C, D, and F, respectively.

Example B.3 (Bernoulli Random Variable)**.** When a random variable is binary then we call it a **Bernoulli** random variable and its probability distrubition is called **Bernoulli** distribution. Consider a random variable that can only take two values, say, 0 or 1. It is common to think of these two values as coding a set criterion with 1 typically assigned if the criterion is met and 0 is assigned for failing to meet the criterion. For example, X could be whether you will get a job right after graduation. If you do then $X = 1$ and if you do not then $X = 0$. Let p denotes the probability that you will get a job. Then, the $p.m.f.$ of the Bernoulli distribution is given by:

$$f(x) = \begin{cases} p & if\ X = 1 \\ 1 - p & if\ X = 0 \end{cases}$$

The $c.d.f$ of the Bernoulli distribution is given by:

$$F(x) = \begin{cases} 0 & if \quad X < 0 \\ 1-p & if \quad 0 \leq X < 1 \\ p & if \quad X \geq 1 \end{cases}$$

B.3.2 Probability distribution of a continuous random variable

In economics a large majority of variables of interest in theory are continuous random variables. For example, the change in the price of Apple stock between two time periods is the return on Apple stock. If you are a trader in the NYSE then the stock return on Apple is a continuous random variable that can take any value on an interval. In such a case we cannot obtain the probability of the random variable taking an exact value. But we can only compute the probability that this random variable will fall in a given interval. So at best we can determine the probability that GDP growth for the US next quarter will be between say 1/% and 2%. This probability is obtained by computing the area under the **probability density function (p.d.f)**. Let X denote a continuous random variable and $f(x)$ denotes the p.d.f. Then,

1. The probability that X takes value over the interval $\{a, b\}$ is given by:

$$P(a \leq X \leq b) = \int_a^b f(x)\ dx$$

2. The c.d.f (the probability that $X \leq x$) is given by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\ dx$$

Below I plot the empirical c.d.f for Apple's stock return. Let X denotes this stock return. From Fig 3.2 we can infer that $P(X \leq 0) = 0.47$ and $P(X \leq 3) = 0.95$.

Figure 3.3 below presents the $p.d.f$ of the daily stock return that corresponds to the $c.d.f$ plotted in Figure 3.2. Using this we can work the probability of stock returns falling in any given interval. For instance, the probability that Apple stock return will fall between 0 and 3% is the area under the p.d.f. between these two values. Figure 3.3 highlights this area and we can see that this probability is equal to 0.47.

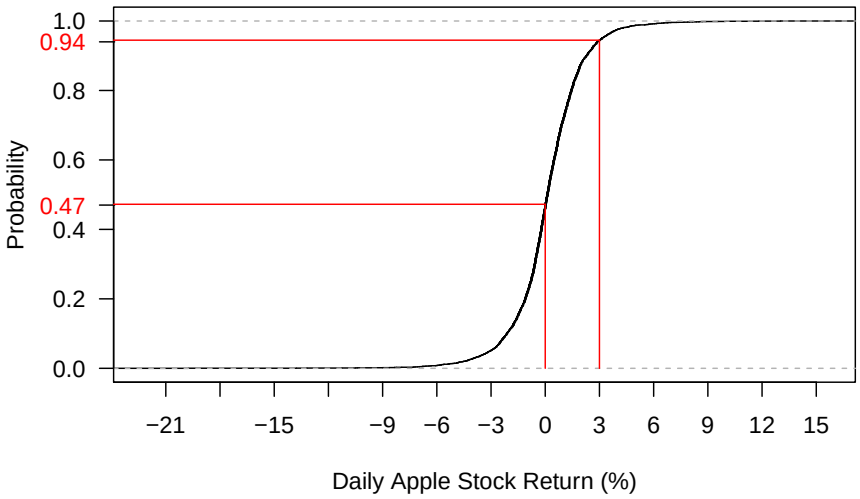


Figure B.2: Empirical c.d.f of daily Apple Stock Return (2007-2019)

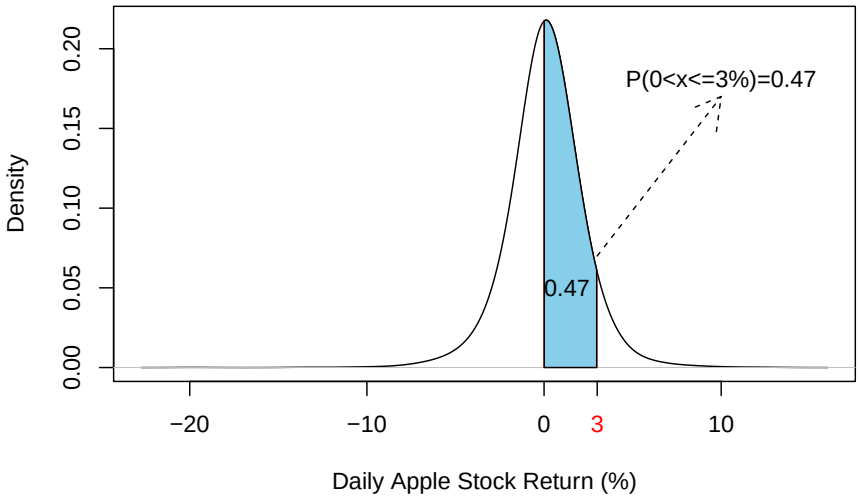


Figure B.3: Empirical p.d.f of daily Apple Stock Return (2007-2019)

B.4 Moments of a probability distribution function

The information contained in a probability distribution can be meaningfully summarized into measures that are called **moments** of that distribution. There are three moments we often use in economics:

1. Center of the distribution: this is first moment of a given probability distribution and it gives us the most likely value of the random variable. It can be measured by mean, median or mode. We will use mean as a measure of the center of the distribution.
2. Width of the distribution: this is the second moment and it measures the average distance from mean under a given probability distribution. We will use standard deviation as a measure of the width of the distribution.
3. Shape of the distribution: this feature relates to role played by **tail events** , i.e., events that have very low probability of happening under a given probability distribution. Two relevant measures are Skewness and Kurtosis

B.4.1 First moment of a probability distribution: Expected value

What is the most likely value of a random variable? To answer that we often compute **expected value** of the random variable which gives us the center (or peak) of the underlying probability distribution. We will use **E** to denote expected value. So $E(X)$ is the expected value of a random variable and we will use μ_X to denote the mean or average value of X.

Definition B.6 (Expected Value)**.** Consider a discrete random variable X that can take n possible values and has the following probabilty distrubution:

X	p(X)
x_1	p_1
x_2	p_2
x_3	p_3
$\vdots$	$\vdots$
x_n	p_n

Then, the expected value of X is given by:

$$E(X) = x_1p_1 + x_2p_2 + ... + x_np_n = \sum_{i=1}^{n} x_ip_i$$

Hence, expected value is a probability-weighted average of all possible values of a random variable. ::: {.example #unnamed-chunk-78} Suppose you toss a fair coin and receive $10 if tails shows and receive 0 if heads shows. What is the expected value of the winnings from a single toss of this coin? :::

Solution. Let X denotes winnings from this game. It can take a value of $10 with a probability of 0.5 and 0 with a probability of half. So the expected value of X is:

$$E(X) = x_1p_1 + x_2p_2 = 10 \times 0.5 + 0 \times 0.5 = \$5$$

Example B.4. Suppose you can invest $10,000 in a mutual fund after 1 year can earn a return of 10% with a probability of 0.1 or a return of 2% with a probability of 0.5 or a loss of 4% with a probability of 0.4. What is the expected return of investing $10,000 in this mutual fund?

Solution. Let X denotes expected return in dollars. It can take 3 possible values: $1000 with a probability of 0.1, $200 with a probability of 0.5, and -$400 with a probability of 0.4 The expected value is given by:

$$E(X) = 1000 \times 0.1 + 200 \times 0.5 - 400 \times 0.4 = \$40$$

As mentioned earlier, the first moment of the probability distribution (i.e., the expected value) gives us the most likely value of the random variable. How useful is this knowledge will depend on how far any realization of the random variable can be from its expected value. The average distance from the average measures the width of the distribution. Wider the distribution, less useful is the knowledge of the expected value.

B.4.2 Second moment of the distribution.

To determine the width or *dispersion* of a probability distribution we use **variance** or **standard deviation**. The variance is the expected value of the squared deviation of each realization of the random variable from its average. We will denote the variance by $Var(X)$ or σ_X^2:

$$Var(x) = \sigma_X^2 = E[(X-\mu_x)^2] = (x_1-\mu_X)^2 \times p_1 + (x_2-\mu_X)^2 \times p_2 + ... + (x_n-\mu_X)^2 \times p_n$$

Or using the summation sign,

$$Var(x) = \sum_{i=1}^{n} (x_i - \mu_x)^2 p_i$$

The standard deviation is simply the square root of the variance and is in the same units as the random variable. This allows easy comparison of the width and the center of the distribution. We will denote standard deviation by σ_X.

Example B.5. Using the mutual fund example, the variance will measure **riskiness** of the investment. It is given by:

$$Var(X) = (1000 - 40)^2 \times 0.1 + (200 - 40)^2 \times 0.5 + (-400 - 40)^2 \times 0.4 = 182400$$

Because variance is in square units and hence hard to interpret, we can easily compute the standard deviation as the square root of the variance:

$$\sigma_X = \sqrt{182400} = \$427.08$$

Hence, even though the average return from this investment is \$40, you can be \$427 above or below this average.

How much can we say about a random variable if we only know its mean and the standard deviation? That depends on the type of distribution the random variable follows. One of the most commonly used distribution in statistic is the **Normal Distribution** or the **Gaussian Distribution**. A random variable that follows normal distribution has a bell-shaped probability distribution with a given mean and standard deviation. One of the most useful features of such a distribution is that knowledge of the first two moments alone is sufficient to characterize the entire probability distribution. Figure 3.4 below shows a normal distribution with a mean of 5 and a standard deviation of 2.

Key features of the normal distribution that are very useful for us:

a. 95% of the values fall within 1.96 times the standard deviation of the mean:

$$P(\mu_X - 1.96\sigma_X \leq X \leq \mu_X + 1.96\sigma_X) = 0.95$$

b. Tail events (low probability events) on either side of the mean are equally unlikely.

c. Central limit theorem: The distribution of sample means calculated from repeated random sampling from a given population approaches a normal distribution as the sample size approaches ∞.

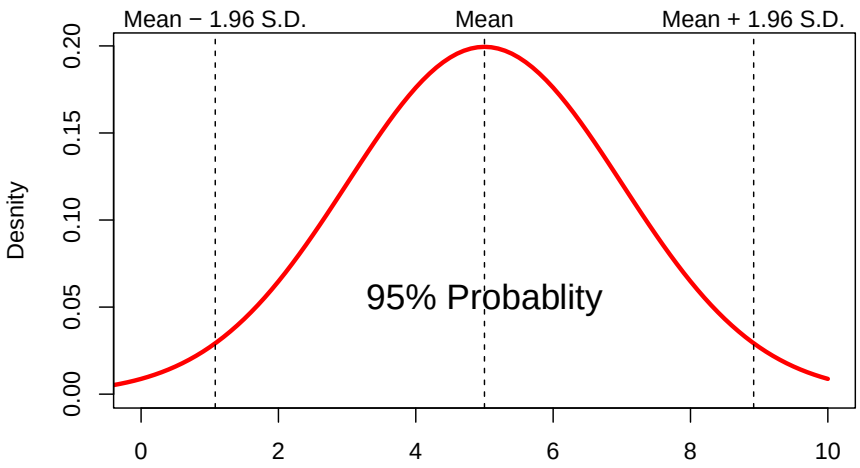


Figure B.4: Normal distribution with mean=5 and s.d.=2

B.4.3 Third and Fourth Moments: Skewness and Kurtosis

In many cases, the distribution of a random variable is not normal and in such cases higher moments provide useful information about the shape of such probability distribution. The shape of the probability distribution plays an important role in many economic and financial applications. There are two measures of shape that are of interest:

1. **Skewness**: this is the third moment of the distribution and it measures how skewed a distribution is. The formula for skewness is given by:

$$Skewness = \frac{E[(X - \mu_X)^3]}{\sigma_X^2}$$

A normal distribution has a skewness of zero. There are two possible types of skewed distributions:

a. A positively skewed distribution will have a long right tail implying lower probability of very large values relative to the mean.

b. A negatively skewed distribution will have a long left tail implying lower probability of very small values relative to the mean.

Figure 3.5 shows three probability distributions. For the left-skewed distribution, a longer left tail indicates low probability of obtaining values below the mean. Similarly, for the right-skewed distribution, a longer right tail indicates low probability of obtaining a value above the mean. For a normal distribution, the probability of obtaining a value above the mean is the same as the probability of obtaining a value below the mean.

2. Kurtosis: this is the fourth moment of the distribution that captures the peakedness of the distribution (or thickness of the tail), i.e., how many observations fall on the extreme ends of a given probability distribution. As a result it tells us the role played by extreme values in driving the variance of a random variable. The formula is given by:

$$Kurtosis = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}$$

A normal distribution has a Kurtosis of 3. A value that is above or below 3 will give us excess or deficient Kurtosis. Two possibilities are:

a. Leptokurtic distribution: has a Kurtosis value greater than three. Such a distribution will have fat tails compared to a normal distribution indicating greater area under the tails.

b. Platykurtic distribution: has a Kurtosis value less than 3. Such a distribution will have thin tails compared to a normal distribution.

Fig 3.6 shows three types of distribution based on their Kurtosis. The leptokurtic distribution has a Kurtosis value of greater than 3 and is more **heavy-tailed** or **peaked** than a normal distribution.

B.5 Useful probability distributions

Using the normal distribution we can derive a few useful probability distributions that are utilized in hypothesis testing.

1. Standard Normal Distribution: A random variable that follows normal distribution with a mean of 0 and standard deviation of 1.

2. Chi-square distribution: is obtained by squaring and adding independent standard normal distribution. For example, is X and Y are two standard normal random variables, then $Z = X^2 + Y^2$ follows a Chi-square distribution with two degrees of freedom.

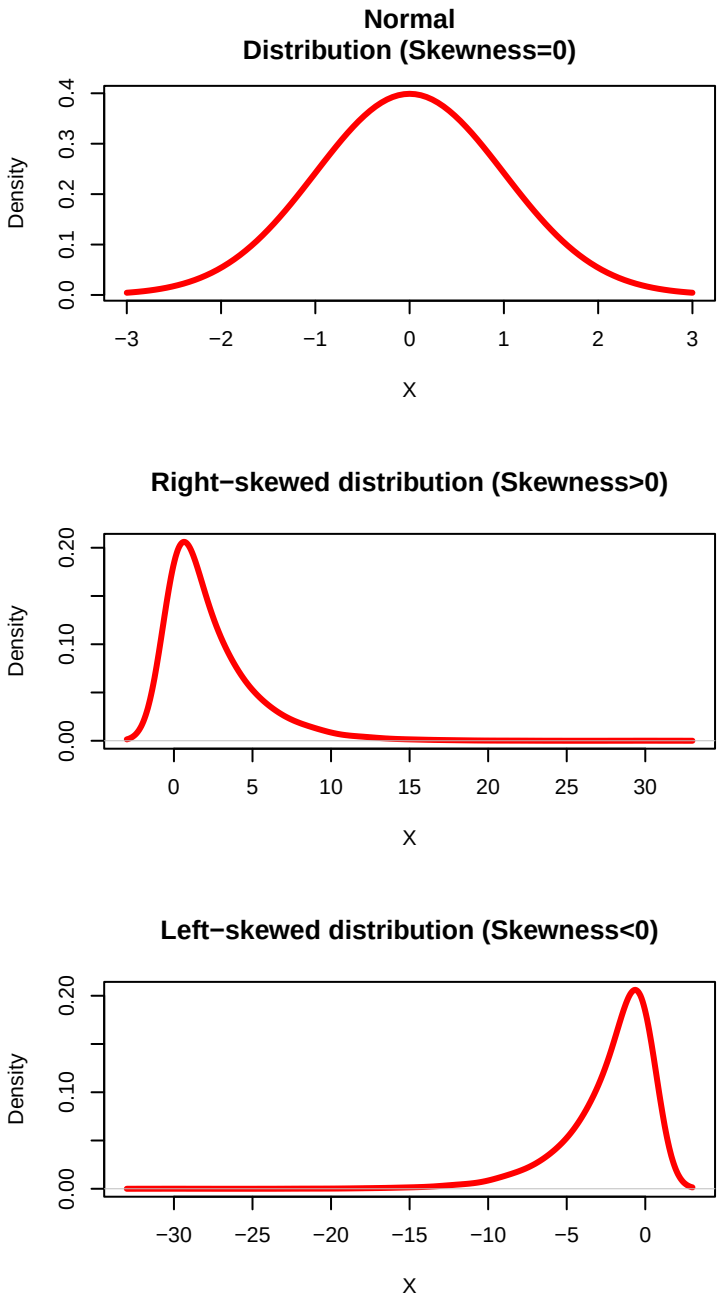


Figure B.5: Skewness of a Probablity distribution

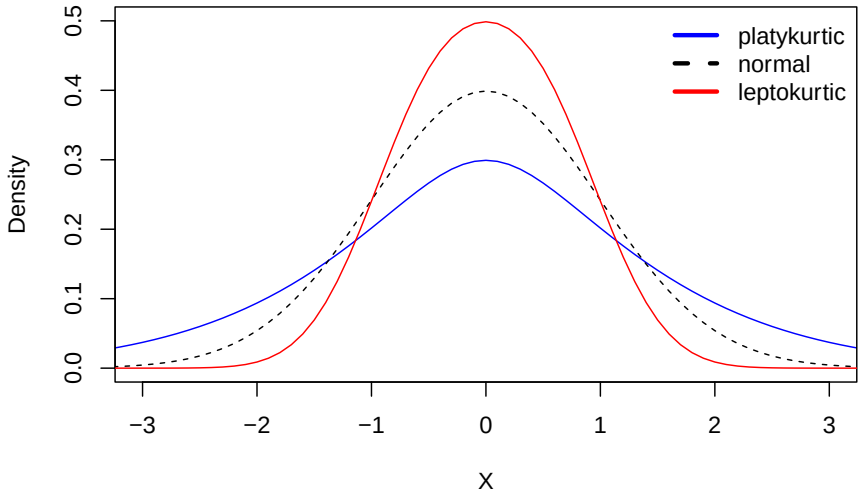
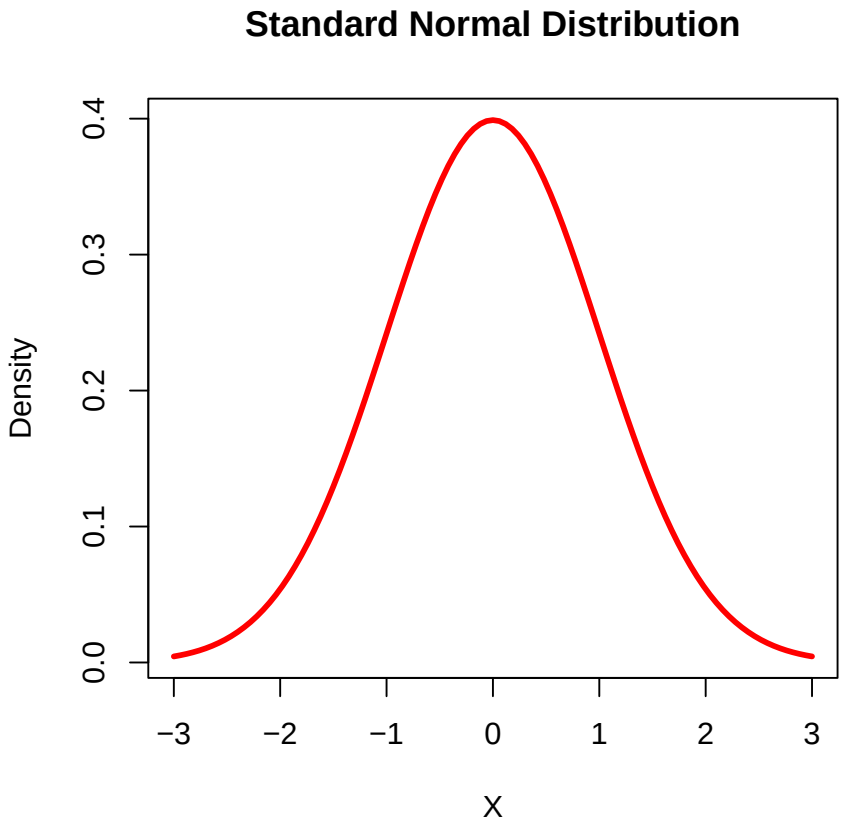
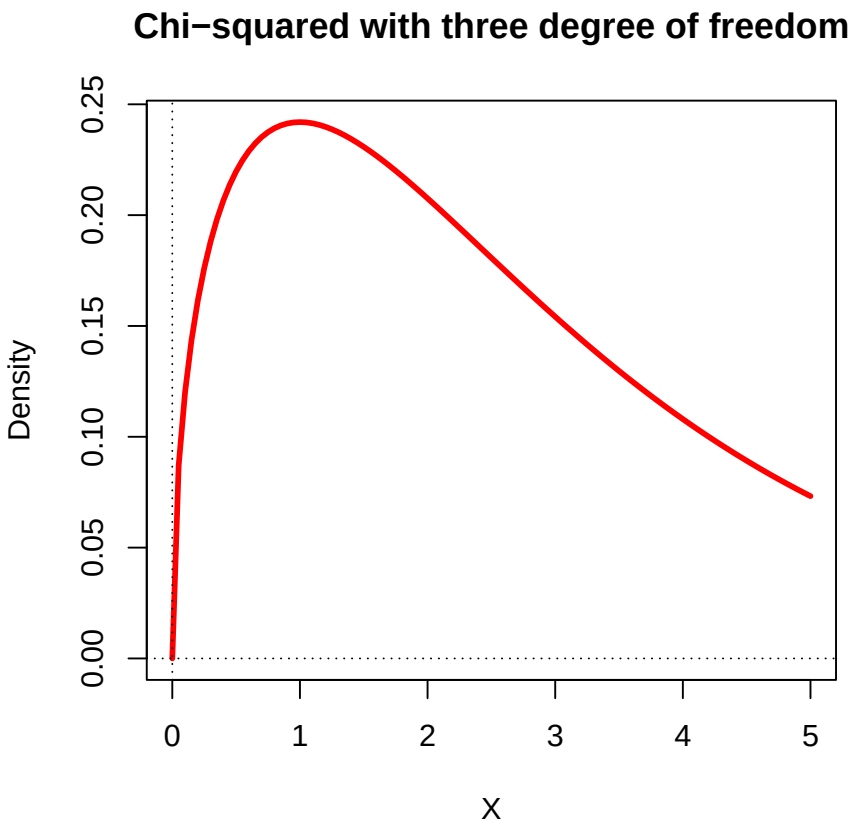


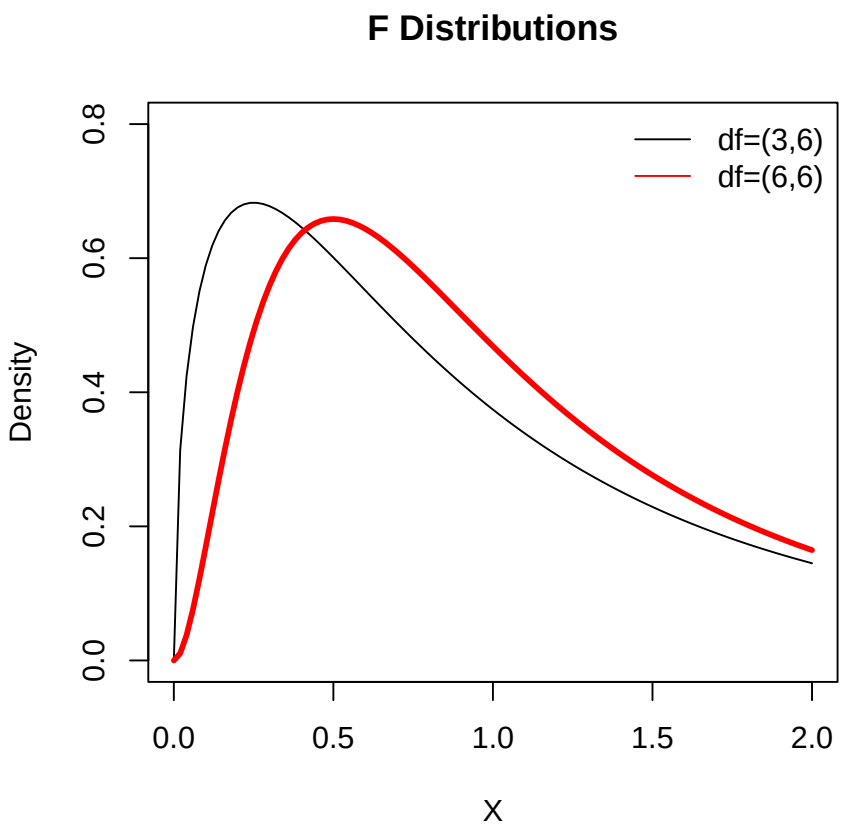
Figure B.6: Kurtosis of a Probablity distribution

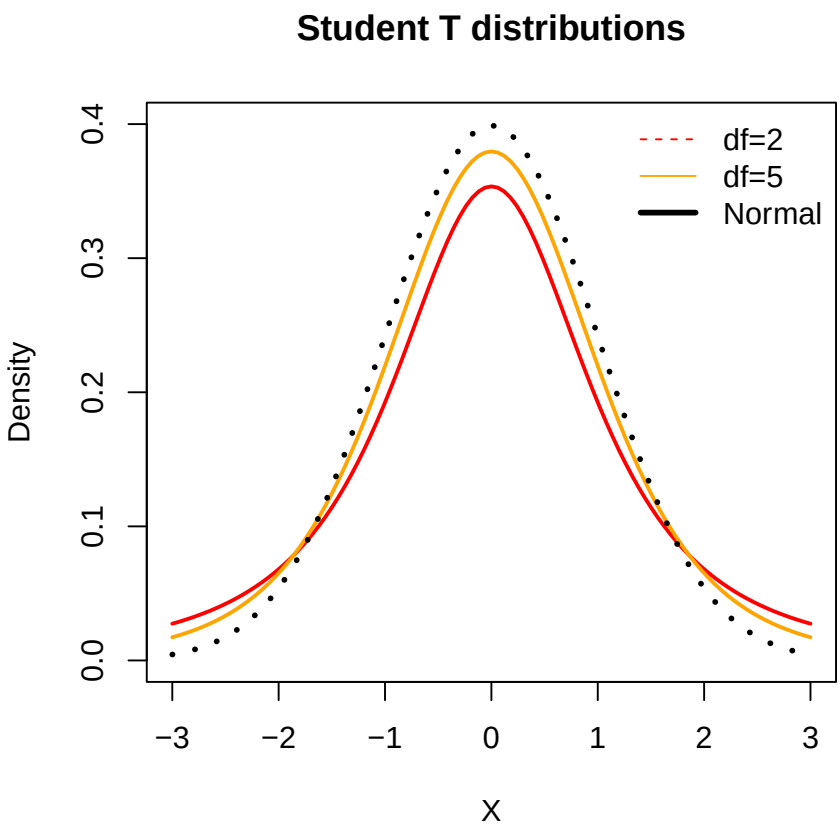
3. F-distribution: is obtained by taking a ratio of two chi-square distribution. For example, if X is Chi-square with v_1 degrees of freedom and Y is a Chi-square with v_2 degrees of freedom, then $Z = \frac{X}{Y}$ follows F-distribution with v_1 and v_2 degrees of freedom.

4. t-distribution: Student's t-distribution is obtained by taking a ratio of a standard normal and the square root of a Chi-square random variable. For example, if X is a standard normal and Y is a Chi-square with m degrees of freedom, then $Z = \frac{X}{\sqrt{Y/m}}$ follows t-distribution with m degrees of freedom. t-distribution has fatter tails when compared to normal.


Standard Normal Distribution
Density
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
X


Chi−squared with three degree of freedom
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
X


F Distributions
df=(3,6)
df=(6,6)
Density
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
X



B.6 Joint Probability Distribution

In economics, often we are interested in the relationship between a pair of variables. For example, how does interest rate affects consumption spending? Or how does education affect wages? In order to statistically answer such questions, we need to understand the meaning of statistical relationship between two or more variables. One way to move forward is to assume that both variables jointly follow some given probability distribution which can be used to infer their relationship with one another.

For simplicity, I will use the discrete random variables case but the concepts covered can be easily extended for the continuous random variables case.

Let X and Y denote two random variables of interest, both from a common probability distribution denoted by $F(x, y)$. This function gives us the proba-

bility that X and Y simultaneously take on certain values:

$$F(x, y) = P(X = x, Y = y)$$

Example B.6. Suppose you are an investment banker and you are considering investment into two assets: a stock listed in NYSE (X) and a cotton futures (Y) listed in Chicago Mercantile Exchange. Suppose X can take three possible values: 2/%, 3/%, or 4/%. Similarly Y can take three possible values given by 6/%,4/%, or 1/%. The value will depend on the state of the economy. Suppose there are three possiblities for the economy next year: boom, expansion, and status quo. The joint probability distribution for X and Y is given by:

State of Economy	X/Y	6	4	1	Total
Recession	2	0.15	0.2	0.1	0.45
Expansion	3	0.1	0.1	0.2	0.4
Status quo	4	0.1	0.05	0	0.15
	Total	0.35	0.35	0.3	1

So in a recession, the probability of obtaining a return of 2/% on the stock and 6/% return on the commodity, i.e, $P(X = 2, Y = 6)$, is 0.15. Using the above joint probability distribution of X and Y we can compute two related distributions for each random variable:

1. Marginal distribution: For each random variable, we can extract its own probability distribution from the joint probability distribution. This is done by simply adding probabilities of all possible outcomes for a particular value of a given random variable. For example, the marginal distribution for X is given by:

$$P(X = x) = \sum_{i=1}^{n} P(X = x, Y = y_i)$$

Hence, in our example, the marginal distribution of X is given by the last column, called Total in the table. For Y it is the row called Total. We can use the marginal distribution to compute the unconditional expected value of each random variable. For example,

$$E(Y) = 6 \times P(Y = 6) + 4 \times P(Y = 4) + 1 \times P(Y = 1) = 3.8$$

2. Conditional distribution: For each random variable, we can also compute its probability distribution conditional on the other variable taking on a specific value. For example,e the conditional distribution of Y given that $X = x$ is given by:

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

From our example, what is the probability of obtaining 4% return on commodity under status quo if the return on the stock is 4%? So here we are interested in finding out:

$$P(Y = 4|X = 4) = \frac{P(X = 4, Y = 4)}{P(X = 4)} = \frac{0.05}{0.15} = 0.33$$

To see this, note that from the table that P(X=4, Y=4) under status quo is given by 0.05. Also, using the definition of marginal distribution, we know that P(X=4)=0.15.

The conditional distribution of a random variable is a first step toward understanding the statistical relationship between two or more random variables. Just like the probability distribution of a random variable has a mean and a variance, the conditional distribution can similarly be characterized by conditional mean and conditional variance:

1. Conditional expected value ($E(Y|X)$): Using the conditional distribution we can now compute the expected value of a random variable, given the value of another random variable. This is denoted by $E(Y|X)$ and can be computed as follows:

$$E(Y|X) = y_1 \times P(Y = y_1|X = x) + y_2 \times P(Y = y_2|X = x) + ... + y_n \times P(Y = y_n|X = x)$$

As we can see, this expected value will be a function of X. Depending on the realization of X our expectation of Y would change. In economics, we can imagine many such examples. For example, given our education level our expected wage will change. Similarly, given expenditure on advertising, expected sales will change. Hence, conditional expected value goes a long way in establishing statistical relationship between economic variables.

Going back to our example, let us compute the expected return on the commodity Y conditional on the information that the return on X is 3%:

$$E(Y|X) = 6 \times P(Y = 6|X = 3) + 4 \times P(Y = 4|X = 3) + 1 \times P(Y = 1|X = 3)$$

Here, $P(Y = 6|X = 3) = \frac{0.1}{0.4} = 0.25$, $P(Y = 4|X = 3) = \frac{0.1}{0.4} = 0.25$ and $P(Y = 1|X = 3) = \frac{0.1}{0.2} = 0.5$. Hence, $E(Y|X = 3) = 3\%$. Contrast this to the unconditional expected value of Y of 3.8% we computed earlier.

2. Conditional variance ($Var(Y|X)$): Now even the variance of a random variable can be affected by another random variable. Here, we are interested in deviations of the random variable from its conditional mean:

$$Var(Y|X) = (y_1 - E(Y|X))^2 \times P(Y = y_1|X = x) + (y_2 - E(Y|X))^2 \times P(Y = y_2|X = x) + ... \\ +(y_n - E(Y|X))^2 \times P(Y = y_n|X = x) \tag{B.3}$$

B.7 Measures of statistical association

We can now define two measures of statistical relationship. The first one is called **Covariance** and the second is **Correlation**.

1. Covariance is a measure of association that captures how deviations from mean of one random variable are related to deviations of another random variable to its respective mean. For example, if your hours of study are above average, then what is your test score relative to average? Formally,

$$Cov(X, Y) = E(Y - \mu_Y)(Y - \mu_X)$$

If the above number is positive, then there is a positive relationship between X and Y. That is, when X is above its mean then Y is also above its mean. If the number is negative then there is a negative relationship between X and Y.

Note that because X and Y are often in different units of measurement, the number we obtain for covariance has no meaning or implication for the strength of the relationship between two variables.

2. Correlation: is the value of covariance that is standardized by dividing this number by standard deviations of each random variable:

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \times \sigma_Y}$$

This number is unit free and falls between -1 and 1. The sign of the correlation tell us about the direction of the relationship whereas the value of the correlation gives information about the strength of the relationship. A higher absolute value indicates stronger statistical relationship between two variables.

B.7.1 Rules of expectation and variances

Here are some useful rules that are useful for our purpose:

1. $E(\beta) = \beta$ and $Var(\beta) = 0$ where β denotes a constant.
2. $E(\beta X) = \beta E(X)$ and $Var(\beta X) = \beta^2 Var(X)$ where β denotes a constant.
3. Consider two random variables X and Y, and let a and b denotes two constants. Then,
   3.1. $E(aX + bY) = aE(X) + bE(Y)$
   3.2. $E(aX - bY) = aE(X) - bE(Y)$
   3.3. $Var(aX+bY) = a^2Var(X)+b^2Var(Y)+2abCor(X,Y)\sqrt{Var(X)}\sqrt{Var(Y)}$
   3.4. $Var(aX-bY) = a^2Var(X)+b^2Var(Y)-2abCor(X,Y)\sqrt{Var(X)}\sqrt{Var(Y)}$

B.8 Sampling and Estimation

An important distinction in statistics is between the population of interest and a sample of this population that we usually work with. Due to feasibility of data collection and cost both in terms of time and money, most real world analysis is based on a sample that is a subset of the population of interest. For example, to study how business major affects starting salary, the relevant population is all business majors from a graduating class in the U.S. in a given year. In practice however, we will most likely use a sample of this population, for example all business majors from JMU. How useful an analysis based on a sample is depends on how representative the chosen sample is of the entire population.

For our purpose, lack of data on population means that the true probability distribution of a random variable is unknown and hence the true values of mean, variance, Covariance etc are also unknown to us. Statistics provides a way of using samples to **estimate** relevant moments of the probability distribution. The approach we take is as follows:

1. Consider the unknown moments of the true probability distribution as ** population parameters** that we would like to estimate.
2. Draw a representative sample from the population. In simple random sampling we draw n observations at random so that each member of the population is equally likely to be included in the sample. We can also use other complex sampling schemes where certain groups of population are more likely to be selected in the sample than others. Two examples:

   a. Suppose we are interested in finding out starting salary of CoB majors at JMU. The population will be every graduating student for a given year. However, we may work with a sample of students, where we draw randomly from every major ensuring that all graduating students have equal probability of selection.
   b. Suppose we are interested in finding out usage of food stamps in Harrisonburg area. The population of interest will be all residents of Harrisonburg who use food stample. However, we may work with a sample where a certain demographic group is more likely to be part of the sample (and hence is *oversampled*).

3. Use the sample to compute sample estimates for each population parameter of interest. For example for expected value we can use sample mean as an estimator, for variance we can use sample variance as an estimator and so on. There are following key differences between population parameters and their sample estimates:

   a. Population parameters are true but unknown values that we are interested in measuring. In contrast, sample estimates can be computed using our sample data.
   b. Population parameters are fixed whereas sample estimates change as we change our sample. For example, if we compute mean starting salary of business majors from JMU we get one number. If use data from UVA we get another number for mean starting salary.
   c. Because different samples give us different sample estimates for the same population parameter, we need to ensure that our sample estimator from one sample data is reliable.

4. Sampling distribution: Hypothetically, we can draw many samples from the same population and compute sample estimate for each sample. This will give us a distribution of for the sample estimate which will have its own mean and variance. We can use this sampling distribution to:

   a. Establish reliablity of the sample estimator. Specifically any sample estimator should be unbiased and efficient. More on this in the next section.
   b. Statsitically test hypotheses about the true population parameter
   ### Unbiasedness and efficiency

Let θ denote a population parameter of interest. For example, it can be the mean of the random variable of interest. Let $\hat{\theta}$ denotes a sample estimator of θ that can be computed using sample data. Then,

1. $\hat{\theta}$ is an **unbiased** estimator of θ if:

$$E(\hat{\theta}) = \theta$$

The idea here is that if we repeatedly draw a sample from the same population and compute $\hat{\theta}$ for each such sample, the average of these estimators must be equal to the true population parameter for unbiasedness. In other words, the center of the sampling distribution is at the true population parameter value.

We can now define **bias** of an estimator as follows:

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$

For an unbiased estimator, $Bias(\hat{\theta}) = 0$. If $Bias(\hat{\theta}) > 0$ then we have an over-estimate and if $Bias(\hat{\theta}) < 0$ then we have an under-estimate.

2. Efficiency: Unbiasedness ensure that the average of sample estimator is equal to the true population parameter. But if the standard deviation of the sample estimator is too high, then knowing that the average is close to the true value is not very useful. In statistics, we call such an estimator unbiased but **imprecise or inefficient**. To be efficient the standard deviation (or variance) of the sample estimator should be as small as possible. Between two unbiased estimators, a more efficient estimator will have a lower variance.

Example B.7. Suppose we have a random sample with n observations: $\{x_1, x_2, ..., x_n\}$ drawn from a population with a mean of μ_x. Sample mean is defined as:

$$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

The expected value of the sample mean is given by:

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{N} x_i}{N}\right)$$

Using properties of the expected value, we get:

$$E(\overline{X}) = \frac{E(x_1) + E(x_2) + ... + E(x_N)}{N}$$

Note that because this is a random sample from the same population with a mean of mu_x, we get $E(x_1) = E(x_2) = .. = E(x_n) = \mu_x$. Hence,

$$E(\overline{X}) = \frac{\overbrace{\mu_x + \mu_x + ... + \mu_x}^{\text{N terms}}}{N} = \mu_x$$

As a result the sample mean is an unbiased estimator of the population mean. However, there are many other possible unbiased estimators of the population mean. We can show that among all other unbiased estimator of the population mean, sample mean has the lowest variance and hence is most efficient estimator as well.

Definition B.7 (Best Unbiased Estimator (BUE))**.** Let θ denote a population parameter of interest. Then, an sample estimator denoted by $\hat{\theta}$ is the best unbiased estimator of θ if the following two conditions are satisfied:

1. $\hat{theta}$ is an unbiased estimator, i.e., $E(\hat{\theta}) = \theta$. In this case the sampling distribution is centered at the true value of the parameter.
2. $\hat{\theta}$ is an efficient estimator, i.e., $Var(\hat{\theta}) < Var(\hat{\theta_A})$ for any other unbiased estimator denoted by $\hat{\theta_A}$. In this case the width of the sampling distribution around the mean is smallest possible.

B.9 Hypothesis testing

An important part of any statistical analysis is testing various hypotheses about population parameters of interest. This is known as *statistical inference* and here we use the sampling distribution of the estimator to formally test whether the corresponding population of interest takes a certain value or not. This is important because even with an unbiased and efficent estimator we do not know the true value of the population parameter of interest. In this section we will look at two types of hypotheses testing procedures that are most relevant for Econometrics. The procedure for any statistical test more or less consists of the following steps:

1. Formulate a hypothesis of interest. This typically manifest as a restriction on the value of a population parameter (or a combination of multiple parameters). The goal is to test whether there is support for this restriction in our sample or not. There are two types of hypotheses that we must formulate:

   1.1. Null Hypothesis (H_0): A null hypothesis is the statement about the population parameter we assume to be true until we find evidence otherwise. For example, we can test whether the population mean of starting salary for CoB majors is \$60,000. Formally,

$$H_0 : \mu_X = 60,000$$

Note that the null hypothesis statement is an equality condition.

1.2. Alternative Hypothesis (H_A): This is the logical counterpart of the null hypothesis and here we specify. There are two types of alternative hypothesis we can specify:

a. Two-sided alternative: Here, the alternative hypothesis statement allows for both sides of the inequality. Going back to our example of starting salary, a two-sided alternative will be:

$$H_A : \mu_X \neq 60,000$$

b. One-sided alternative: Here, we either use a greater or less than sign for the alternative hypothesis. So for example, we can speficy the following one-sided alternative:

$$H_A : \mu_X > 60,000$$

2. Compute the relevant test statistic that is a function of the sample data. The formula for the test statistic is a function of the sample estimator and the value of the population parameter(s) we assumed in the null hypothesis.

3. The test statistic is assumed to follow a certain probability distribution under the assumption that the null hypothesis is correct. The tails of this distribution summarizes values of the test statistic that are less likely to realize. Such a value of the test statistic provides us a threshold level, called the **critical value**, beyond which the test statistic values are less likely to realize if our hypothesis is true. The decision rule for rejecting or not rejecting the null hypothesis is based on the comparison between the computed test statistic and the associated critical value.

Note that there is always a measure of uncertainty in any hypothesis testing: we may end up making a wrong decision. There are two types of errors we can make here:

1. **Type I** error: here we reject H_0 when it is true. The probability of this type of error is denoted by α and is called the **level of significance** of a test.

2. **Type II** error: here we do not reject H_0 when it is false. The probability of this type of error is related to the **power** of a test.

Ideally we would like to minimize the probability of both types of errors but we cannot do that because reducing one error comes at the cost of increasing the other. As a result, we first specify an **acceptable** level of significance (type one error probability) and then try to minimize the probability of type two error (or

maximize the power of the test). It is common to assume a level of significance of 5% or $\alpha = 0.05$. So here we are willing to tolerate a 5% chance of falsely rejecting the null hypothesis.

Once we have fixed the level of significance, we can use the distribution table of the test-statistic to obtain the corresponding critical value(s).

B.9.1 Testing a restriction on a single population parameter

Here our goal is to develop tests for testing statements about a single population parameter of interest. So for example, we can either test a statement about a population mean or a population variance.

Example B.8 (t-test for population mean)**.** Suppose you are interested in measuring mean hourly wage of males aged 25-35. Accordingly, we collect a sample of 100 workers from the population of male in this age group with a mean of μ_X and a standard deviation of σ_X. The sample mean is $\hat{\mu}_X = \$25$ and the sample standard deviation is $\hat{\sigma}_X = \$7$. Now, suppose we want to test the following hypothesis:

$$H_0 : \mu_X = 27$$

$$H_0 : \mu_X \neq 27$$

The test statistic is given by the *t-statistic* where:

$$t = \frac{\hat{\mu}_X - \mu_X}{s.e.(\hat{\mu}_X)}$$

where $s.e.(\hat{\mu}_X) = \dfrac{\hat{\sigma}_X}{\sqrt{N}}$ is the standard error of sample mean and N denotes sample size.

If the null hypothesis is true, this test statistic follows **t-distribution** with N-1 degrees of freedom. Using the t-distribution table we can then compute the critical value which is used in formulating the decision rule. Let t_c denote this critical value from the distribution table. Then,

$$|t| > t_c \quad \Rightarrow \text{reject } H_0$$

$$|t| < t_c \quad \Rightarrow \text{do not reject } H_0$$

In our example, $N = 100$, and

$$t = \frac{25 - 27}{\frac{7}{\sqrt{100}}} = -2.86$$

The degrees of freedom is $N - 1 = 99$ and at 5% level of significance the critical value from the t-distribution table is $t_c = 1.98$. Because $|t|$ is larger than the critical value, we reject the null hypothesis. Hence, we find evidence against the statement that the mean hourly wage of male workers is \$25.

Note that an alternative way of testing hypothesis like this is to use the **p-value** rule. The underlying idea is to find out the largest significance level at which we will fail to reject the null hypothesis. This value is called the p-value and most statistical softwares report this value. The decision-rule is then greatly simplified:

If p-value is less than the chosen level of significance (value of α) then reject H_0.

In our case, the p-value is 0.0053. Because we chose $\alpha = 0.05$, according to the p-value rule we will reject the null hypothesis.

Example B.9 (Chi-square test for population variance)**.** Using the same example, we also test a statement about the population variance. Suppose we want to test whether the variance of the hourly wage is 52.

$$H_0 : \sigma_X^2 = 52$$

$$H_0 : \sigma_X^2 > 52$$

The test statistic is given by the *V-statistic* where:

$$V = \frac{(N-1) \times \hat{\sigma_X^2}}{\sigma_X^2}$$

If the null hypothesis is true, this test statistic follows **Chi-square distribution** with N-1 degrees of freedom. Using the distribution table we can then compute the critical value which is used in formulating the decision rule. Let V_c denote this critical value from the distribution table. Then,

$$V > V_c \quad \Rightarrow \text{reject } H_0$$

$$V < V_c \quad \Rightarrow \text{do not reject } H_0$$

In our example,

$$V = \frac{(100-1) \times 7^2}{52} = 93.29$$

The degrees of freedom is $N-1 = 99$ and at 5% level of significance the critical value from the Chi-square distribution table is $V_c = 124.34$. Because V is smaller than the critical value, we do not reject the null hypothesis. Hence, we find no sample Bevidence against the statement that the variance of the hourly wage of male workers is 52.

B.9.2 Testing a restriction on multiple population parameter

Often we are interested in testing a restriction that is a linear combination of two or more population means. Similarly, we maybe interested in comparing the variance of two different populations. In such cases we need to develop statistical tests that allow for comparison between parameters of different populations with given means and variances.

Example B.10 (t-test for comparing population mean of two populations)**.** Suppose you are interested in comparing mean weekly hours studied by Econ majors (X) and non-Econ majors in the college of business. For this purpose, you collect a sample of 25 econ majors and a sample of 30 non-econ majors. The sample mean of weekly hours studied by econ majors is 10 hours with a standard deviation of 4 hours. The sample mean of weekly hours studied by non-econ majors is 8 hours with a standard deviation of 2 hours. Test whether mean weekly hours studied by econ majors is more than the mean weekly hours studied by non-Econ majors.

Let X denote hours studied, N_X denotes sample size, $\hat{\mu}_X$, and $\hat{\sigma}_X$ denote sample mean and standard deviation, respectively for econ majors. Similarly, let Y denote hours studied, N_Y denotes sample size, $\hat{\mu}_Y$, and $\hat{\sigma}_Y$ denote sample mean and standard deviation, respectively for non-econ majors.

The first step, as usual, is to formulate the null and the alternative hypotheses:

$$H_0 = \mu_X - \mu_Y = 0$$
$$H_A = \mu_X - \mu_Y > 0$$

The next step is to compute the relevant test statistic, which in this case is the t-ratio given by:

$$t = \frac{(\widehat{\mu_X} - \widehat{\mu_Y}) - 0}{s.e.(\widehat{\mu_X} - \widehat{\mu_Y})}$$

Using the properties of variance and assuming that two samples are independent of each other, we get:

$$s.e.(\hat{\mu_X} - \hat{\mu_Y}) = \sqrt{Var(\hat{\mu_X}) + Var(\hat{\mu_Y})}$$

Note that $Var(\hat{\mu_X}) = \frac{\widehat{\sigma_X}^2}{N_X}$ and $Var(\hat{\mu_Y}) = \frac{\widehat{\sigma_Y}^2}{N_Y}$. Using this we get:

$$s.e.(\hat{\mu_X} - \hat{\mu_Y}) = \sqrt{\frac{\widehat{\sigma_X}^2}{N_X} + \frac{\widehat{\sigma_Y}^2}{N_Y}} = \sqrt{\frac{16}{25} + \frac{4}{30}} = 0.88$$

So, $t = \frac{10-8}{0.88} = 2.27$

The sample size here is $N_X + N_Y = 55$. Using 5% level of significance and degrees of freedom of 53, the critical value from the t-distribution table for the one-sided alternative is 1.67. Because the |t| is more than 1.67, we reject the null hypothesis. We find evidence for econ majors studying more on average than non-econ majors in our sample.

Example B.11 (F-test for comparing population variance of two populations)**.** Often we may be interested in comparing the variability between two populations. Using our previous example, we may want to test whether variability in hours studied is bigger for econ majors versus non-econ majors. This can be tested by comparing the ratio of two variances against the value of 1. As before, we start by formulating the null and the alternative hypotheses:

$$H_0 : \sigma_X^2/\sigma_Y^2 = 1$$

$$H_0 : \sigma_X^2/\sigma_Y^2 > 1$$

The corresponding test statistic is the F-ratio:

$$F = \frac{\hat{\sigma_X^2}}{\hat{\sigma_Y^2}} = \frac{4^2}{2^2} = 4$$

If the null hypothesis is true, the above test statistic follows F-distribution with $N_x - 1$ degrees of freedom for the numerator and $N_y - 1$ degrees of freedom for the denominator. At 5% level of significance, the critical value for $\nu_1 = 24$ and $\nu_2 = 29$ from the F-distribution table is 3. Because the computed F-ratio exceeds the critical value we reject the null hypothesis.

B.9.3 Confidence interval and Hypothesis testing

One issue with using a sample to estimate population parameters is that by definition a sample estimator will be different for different samples. Thus, sample mean provides no information about how close this estimator is to the true population mean. This uncertainty in estimation can be summarized by computing the standard deviation, with higher value of standard deviation indicating greater uncertainty about the true population parameter. A better measure of this uncertainty is the **confidence interval**.

Definition B.8 (Confidence Interval)**.** Suppose we draw a random sample $\{x_1, x_2, ..., x_N\}$ from a normally distributed population with mean of μ_X and a standard deviation of σ_X. Let $\hat{\mu_X}$ denotes the sample mean and $\hat{\sigma_X}$ denotes sample standard deviation. Then, the 95% confidence interval for $\hat{\mu_X}$ is given by:

$$\left[\hat{\mu_X} - t_{c,2-sided} \times \frac{\hat{\sigma_X}}{\sqrt{N}}, \hat{\mu_X} + t_{c,2-sided} \times \frac{\hat{\sigma_X}}{\sqrt{N}}\right]$$

where $t_{c,2-sided}$ is the critical value that can be obtained from the t-distribution table for a given level of signicance and degrees of freedom. For example, for a 95% confidence interval we will use 5% level of significance.

Example B.12. Suppose N=20, $\hat{\mu_X} = 5$, and $\hat{\sigma_X} = 2$. Then, the 95% confidence interval for $\hat{\mu_X}$ is given by:

$$\left[5 - 2.093 \times \frac{2}{\sqrt{20}}, 5 + 2.093 \times \frac{2}{\sqrt{20}}\right] = [4.06, 5.94]$$

Hence, before we drew our sample from the population, there is a 95% chance that the true population parameter (μ_X) will fall between 4.12 and 5.94. Note that:

1. Wider the confidence interval, greater is the uncertainty about the true value of the population mean.
2. We can use the confidence interval to conduct hypothesis testing for a **two-sided** alternative hypothesis. If the null hypothesis value does not fall in the confidence interval, then with 95% confidence (or at 5% level of significance) we can reject the null hypothesis. For example, consider the following test:

$$H_0 : \mu_X = 3.8$$
$$H_A : \mu_X \neq 3.8$$

Because 3.8 is not in the confidence interval we will reject the null hypothesis at 5% level of significance. Note that we will obtain the same conclusion if we were to compute the t-ratio and compare it with the corresponding critical value from the t-distribution table.

Problems

Exercise B.1. Suppose you roll a 6-sided fair dice. If an odd number shows you win $10. If either 2 or 4 shows you lose $5. If 6 shows, you neither gain nor lose anything.

a. Denote the winnings from this game as X. Tabulate the probability distribiution of the random variable X.

b. Compute the expected value and the standard deviation for X.

Exercise B.2. Consider a population with a mean of μ and variance of σ^2. Suppose you draw a random sample $X_1, X_2, .., X_N$.

a. Show that $\widehat{\mu_A} = 0.25 \times X_1 + 0.25 \times X_3 + 0.25 \times X_8 + 0.25X_{20}$ is an unbiased estimator of μ.

b. Show that $\widehat{\mu_B} = 0.1 \times X_1 + 0.1 \times X_3 + 0.5 \times X_8 + 0.3 \times X_{11}$ is an unbiased estimator of μ.

c. Now compute variance of $\widehat{\mu_A}$ and $\widehat{\mu_B}$. Which one is more efficient estimator of μ.

Exercise B.3. Suppose you collect a random sample of 100 observations and find that sample mean is -25 and sample variance is 350.

a. Test whether the population mean is -22.

b. Test whether the population variance is 400.

Exercise B.4. Suppose you are interested in comparing performance of two different mutual funds, X and Y. Let μ_X and μ_Y denote unknown population mean returns on investment in X and Y, respectively. Suppose you collect past 20 months data for both mutual funds and find that sample mean for fund X is 2% with a standard deviation of 0.5%. In contrast, the sample mean for fund Y is 5% with a standard deviation of 2%.

a. Test whether mean return on Y is greater than that on X.

b. Test whether variance of Y is greater than that of X.

c. Compute the 95% confidence interval for $\widehat{\mu_X}$. Using the confidence interval, what can you say about the population mean return for fund X?

Solutions

Exercise 2.1:

a. The probability distribution is given by:

X	P(X)
10	$\frac{1}{2}$
-5	$\frac{1}{3}$
0	$\frac{1}{6}$

b. The expected value is given by:

$$E(X) \equiv \mu_X = 10 \times \frac{1}{2} - 5 \times \frac{1}{3} + 0 \times \frac{1}{6} = \$3.33$$

The variance is given by:

$$Var(X) \equiv \sigma_X^2 = (10 - 3.33)^2 \times \frac{1}{2} + (-5 - 3.33)^2 \times \frac{1}{3} + (0 - 3.33)^2 \times \frac{1}{6} = 47.22$$

The standard deviation is the square root of the variance: \$6.87.

Exercise 2.2:

a.

$$E(\widehat{\mu_A}) = 0.25 \times E(X_1) + 0.25 \times E(X_3) + 0.25 \times E(X_8) + 0.25E(X_{20})$$

Because, each realization of X is drawn from the same population, they have the same mean giving us:

$$E(\widehat{\mu_A}) = 0.25 \times \mu + 0.25 \times \mu + 0.25 \times \mu + 0.25\mu = \mu$$

Hence, $\widehat{\mu_A}$ is an unbiased estimator of μ.

b.

$$E(\widehat{\mu_B}) = 0.1 \times E(X_1) + 0.1 \times E(X_3) + 0.5 \times E(X_8) + 0.3 \times E(X_{11})$$

Again, because each realization of X has the same mean we get:

$$E(\widehat{\mu_B}) = 0.1 \times \mu + 0.1 \times \mu + 0.5 \times \mu + 0.3 \times \mu = \mu$$

Hence, $\widehat{\mu_B}$ is an unbiased estimator of μ.

c. Before computing the variance, note that each value of X is independent from any other value, implying there is no covariance between any two values of X. As a result we get:

$$Var(\widehat{\mu_A}) = 0.25^2 \times Var(X_1) + 0.25^2 \times Var(X_3) + 0.25^2 \times Var(X_8) + 0.25^2 \times Var(X_{20})$$

Each X has the same variance, giving us:

$$Var(\widehat{\mu_A}) = 0.0625 \times \sigma^2 + 0.0625 \times \sigma^2 + 0.0625 \times \sigma^2 + 0.0625 \times \sigma^2 = 0.25\sigma^2$$

Similarly,

$$Var(\widehat{\mu_B}) = 0.1^2 \times \sigma^2 + 0.1^2 \times \sigma^2 + 0.5^2 \times \sigma^2 + 0.3^2 \times \sigma^2 = 0.36\sigma^2$$

More efficicent estimator has a smaller variance. Based on our calculations, $\widehat{\mu_A}$ has a lower variance and hence is more efficient.

Exercise 2.3:

Suppose you collect a random sample of 100 observations and find that sample mean is -25 and sample variance is 350.

a. The null and alternative hypotheses are given by:

$$H_0 : \mu = -22$$
$$H_A : \mu \neq -22$$

The t-statistic is given by $t = \dfrac{-25 - (-22)}{\sqrt{350}/\sqrt{100}} = -1.603$

The critical value from the t-distribution table at 5% level of significance, 2-tailed, and degrees of freedom of 99 is 1.984. Because $|t| = 1.603 < 1.984$, we do not reject the null hypothesis. Hence, we do not find sufficient evidence in our sample against the null hypothesis.

b. The null and alternative hypotheses are given by:

$$H_0 : \sigma^2 = 400$$
$$H_A : \text{Not } H_0$$

The V-statistic is given by $V = (100 - 1) \times \dfrac{350}{400} = 86.25$

The critical value from the Chi-square distribution table at 5% level of significance and 99 degrees of freedom is given by 124.34. Because the V-statistic value is less than the critical value, we do not reject the null hypothesis.

Exercise 2.4:

a. The null and alternative hypotheses are given by:

$$H_0 : \mu_Y - \mu_X = 0$$
$$H_A : \mu_Y - \mu_X > 0$$

The t-statistic is given by $t = \dfrac{(5-2)-0}{\sqrt{\frac{0.5^2}{20} + \frac{2^2}{20}}} = 6.51$

The critical value from the t-distribution table at 5% level of significance, one-tailed, and 38 degrees of freedom is 1.684. Because the test statistic is greater than the critical value, we reject the null hypothesis. Hence, in our sample we find sufficient evidence against the null hypothesis and conclude that mean return on asset Y is greater than that on asset X.

b. The null and alternative hypotheses are given by:

$$H_0 : \sigma_Y^2 = \sigma_X^2$$
$$H_A : \sigma_Y^2 > \sigma_X^2$$

The F-statistic is given by: $F = \dfrac{2^2}{0.5^2} = 16$. The critical value from the F-distribution with 5% level of significance, and 19 degrees of freedom for both the numerator and the denominator is 2.38. Because the F-statistic is bigger than the critical value, we reject the null hypothesis.

c. The 95% confidence interval for $\widehat{\mu_X}$ is given by:

$$\widehat{\mu_X} \pm \frac{\widehat{\sigma_X}}{\sqrt{N}} \times t_{c,2-sided} = 2 \pm \frac{0.5}{\sqrt{20}} * 2.093 = \{1.766, 2.234\}$$

Hence, before we drew our sample from the population, there is a 95% chance that the true population parameter (μ_X) will fall between 1.766 and 2.234. Further, because zero is not inlcuded in the confidence band, we can say with 95% confidence that the mean return on asset X is statistically significant.

In macroeconomics, the above relationship is known as the **Keynesian** consumption function. John Maynard Keynes proposed that at the aggregate level, consumption changes in proportion to changes in disposable income:

$$C = a + b\ Y^D$$

Here C denotes private consumption expenditure and Y^D denotes post-tax or disposable income. a is the intercept and captures the part of consumption that is indpenedent of income. b is the slope and measures the unit change in consumption caused by a unit change in disposable income. b measures the marginal propensity to consume and is a parameter of interest we would like to estimate using data.

Appendix C

Statistical Tables

Table A: Critical Values for the t-distribution

1-tailed2-tailed	0.100.20	0.050.1	0.0250.05	0.010.02	0.0050.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797

1-tailed2-tailed	0.100.20	0.050.1	0.0250.05	0.010.02	0.0050.01
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
70	1.294	1.667	1.994	2.381	2.648
80	1.292	1.664	1.990	2.374	2.639
90	1.291	1.662	1.987	2.368	2.632
100	1.290	1.660	1.984	2.364	2.626
110	1.289	1.659	1.982	2.361	2.621
120	1.289	1.658	1.980	2.358	2.617
Inf	1.282	1.645	1.960	2.326	2.576

Table B: Critical Values for the Chi-square distribution

df	0.10	0.05	0.01
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21
11	17.28	19.68	24.72
12	18.55	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.59	33.41
18	25.99	28.87	34.81
19	27.20	30.14	36.19

df	0.10	0.05	0.01
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.42	42.98
25	34.38	37.65	44.31
26	35.56	38.89	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59
30	40.26	43.77	50.89
40	51.81	55.76	63.69
50	63.17	67.50	76.15
60	74.40	79.08	88.38
70	85.53	90.53	100.43
80	96.58	101.88	112.33
90	107.57	113.15	124.12
100	118.50	124.34	135.81
110	129.39	135.48	147.41
120	140.23	146.57	158.95

Table C: 1% Critical Values for the F distribution

df2/df1	1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09

df2/df1	1	2	3	4	5	6	7	8	9	10
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
31	7.53	5.36	4.48	3.99	3.67	3.45	3.28	3.15	3.04	2.96
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93
33	7.47	5.31	4.44	3.95	3.63	3.41	3.24	3.11	3.00	2.91
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86
37	7.37	5.23	4.36	3.87	3.56	3.33	3.17	3.04	2.93	2.84
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83
39	7.33	5.19	4.33	3.84	3.53	3.30	3.14	3.01	2.90	2.81
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
41	7.30	5.16	4.30	3.81	3.50	3.28	3.11	2.98	2.87	2.79
42	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78
43	7.26	5.14	4.27	3.79	3.48	3.25	3.09	2.96	2.85	2.76
44	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75
45	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74
46	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73
47	7.21	5.09	4.23	3.75	3.43	3.21	3.05	2.92	2.81	2.72
48	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.71
49	7.18	5.07	4.21	3.73	3.42	3.19	3.03	2.90	2.79	2.71
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70
51	7.16	5.05	4.19	3.71	3.40	3.18	3.01	2.88	2.78	2.69
52	7.15	5.04	4.18	3.70	3.39	3.17	3.00	2.87	2.77	2.68
53	7.14	5.03	4.17	3.70	3.38	3.16	3.00	2.87	2.76	2.68
54	7.13	5.02	4.17	3.69	3.38	3.16	2.99	2.86	2.76	2.67
55	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66
56	7.11	5.01	4.15	3.67	3.36	3.14	2.98	2.85	2.74	2.66
57	7.10	5.00	4.15	3.67	3.36	3.14	2.97	2.84	2.74	2.65
58	7.09	4.99	4.14	3.66	3.35	3.13	2.96	2.83	2.73	2.64
59	7.08	4.98	4.13	3.65	3.34	3.12	2.96	2.83	2.72	2.64
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
61	7.07	4.97	4.12	3.64	3.33	3.11	2.95	2.82	2.71	2.63
62	7.06	4.96	4.11	3.64	3.33	3.11	2.94	2.81	2.71	2.62
63	7.06	4.96	4.11	3.63	3.32	3.10	2.94	2.81	2.70	2.62
64	7.05	4.95	4.10	3.63	3.32	3.10	2.93	2.80	2.70	2.61
65	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61
66	7.04	4.94	4.09	3.62	3.31	3.09	2.92	2.79	2.69	2.60
67	7.03	4.94	4.09	3.61	3.30	3.08	2.92	2.79	2.68	2.60
68	7.02	4.93	4.08	3.61	3.30	3.08	2.91	2.78	2.68	2.59
69	7.02	4.93	4.08	3.60	3.29	3.08	2.91	2.78	2.68	2.59
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59

df2/df1	1	2	3	4	5	6	7	8	9	10
71	7.01	4.92	4.07	3.60	3.29	3.07	2.90	2.77	2.67	2.58
72	7.00	4.91	4.07	3.59	3.28	3.06	2.90	2.77	2.66	2.58
73	7.00	4.91	4.06	3.59	3.28	3.06	2.89	2.77	2.66	2.57
74	6.99	4.90	4.06	3.58	3.28	3.06	2.89	2.76	2.66	2.57
75	6.99	4.90	4.05	3.58	3.27	3.05	2.89	2.76	2.65	2.57
76	6.98	4.90	4.05	3.58	3.27	3.05	2.88	2.75	2.65	2.56
77	6.98	4.89	4.05	3.57	3.26	3.05	2.88	2.75	2.65	2.56
78	6.97	4.89	4.04	3.57	3.26	3.04	2.88	2.75	2.64	2.56
79	6.97	4.88	4.04	3.57	3.26	3.04	2.87	2.75	2.64	2.55
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55
81	6.96	4.88	4.03	3.56	3.25	3.03	2.87	2.74	2.63	2.55
82	6.95	4.87	4.03	3.56	3.25	3.03	2.87	2.74	2.63	2.54
83	6.95	4.87	4.03	3.55	3.25	3.03	2.86	2.73	2.63	2.54
84	6.95	4.87	4.02	3.55	3.24	3.02	2.86	2.73	2.63	2.54
85	6.94	4.86	4.02	3.55	3.24	3.02	2.86	2.73	2.62	2.54
86	6.94	4.86	4.02	3.55	3.24	3.02	2.85	2.73	2.62	2.53
87	6.94	4.86	4.02	3.54	3.24	3.02	2.85	2.72	2.62	2.53
88	6.93	4.85	4.01	3.54	3.23	3.01	2.85	2.72	2.62	2.53
89	6.93	4.85	4.01	3.54	3.23	3.01	2.85	2.72	2.61	2.53
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52
91	6.92	4.85	4.00	3.53	3.23	3.01	2.84	2.71	2.61	2.52
92	6.92	4.84	4.00	3.53	3.22	3.00	2.84	2.71	2.61	2.52
93	6.92	4.84	4.00	3.53	3.22	3.00	2.84	2.71	2.60	2.52
94	6.91	4.84	4.00	3.53	3.22	3.00	2.84	2.71	2.60	2.52
95	6.91	4.84	3.99	3.52	3.22	3.00	2.83	2.70	2.60	2.51
96	6.91	4.83	3.99	3.52	3.21	3.00	2.83	2.70	2.60	2.51
97	6.90	4.83	3.99	3.52	3.21	2.99	2.83	2.70	2.60	2.51
98	6.90	4.83	3.99	3.52	3.21	2.99	2.83	2.70	2.59	2.51
99	6.90	4.83	3.99	3.51	3.21	2.99	2.83	2.70	2.59	2.51
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
Inf	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Table D: 5% Critical Values for the F distribution

df2/df1	1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54

df2/df1	1	2	3	4	5	6	7	8	9	10
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14
33	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11
37	4.11	3.25	2.86	2.63	2.47	2.36	2.27	2.20	2.14	2.10
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09
39	4.09	3.24	2.85	2.61	2.46	2.34	2.26	2.19	2.13	2.08
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
41	4.08	3.23	2.83	2.60	2.44	2.33	2.24	2.17	2.12	2.07
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06
43	4.07	3.21	2.82	2.59	2.43	2.32	2.23	2.16	2.11	2.06
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04
47	4.05	3.20	2.80	2.57	2.41	2.30	2.21	2.14	2.09	2.04
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03
49	4.04	3.19	2.79	2.56	2.40	2.29	2.20	2.13	2.08	2.03
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
51	4.03	3.18	2.79	2.55	2.40	2.28	2.20	2.13	2.07	2.02
52	4.03	3.18	2.78	2.55	2.39	2.28	2.19	2.12	2.07	2.02
53	4.02	3.17	2.78	2.55	2.39	2.28	2.19	2.12	2.06	2.01
54	4.02	3.17	2.78	2.54	2.39	2.27	2.18	2.12	2.06	2.01
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01
56	4.01	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.05	2.00
57	4.01	3.16	2.77	2.53	2.38	2.26	2.18	2.11	2.05	2.00
58	4.01	3.16	2.76	2.53	2.37	2.26	2.17	2.10	2.05	2.00
59	4.00	3.15	2.76	2.53	2.37	2.26	2.17	2.10	2.04	2.00

df2/df1	1	2	3	4	5	6	7	8	9	10
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
61	4.00	3.15	2.76	2.52	2.37	2.25	2.16	2.09	2.04	1.99
62	4.00	3.15	2.75	2.52	2.36	2.25	2.16	2.09	2.03	1.99
63	3.99	3.14	2.75	2.52	2.36	2.25	2.16	2.09	2.03	1.98
64	3.99	3.14	2.75	2.52	2.36	2.24	2.16	2.09	2.03	1.98
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98
66	3.99	3.14	2.74	2.51	2.35	2.24	2.15	2.08	2.03	1.98
67	3.98	3.13	2.74	2.51	2.35	2.24	2.15	2.08	2.02	1.98
68	3.98	3.13	2.74	2.51	2.35	2.24	2.15	2.08	2.02	1.97
69	3.98	3.13	2.74	2.50	2.35	2.23	2.15	2.08	2.02	1.97
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97
71	3.98	3.13	2.73	2.50	2.34	2.23	2.14	2.07	2.01	1.97
72	3.97	3.12	2.73	2.50	2.34	2.23	2.14	2.07	2.01	1.96
73	3.97	3.12	2.73	2.50	2.34	2.23	2.14	2.07	2.01	1.96
74	3.97	3.12	2.73	2.50	2.34	2.22	2.14	2.07	2.01	1.96
75	3.97	3.12	2.73	2.49	2.34	2.22	2.13	2.06	2.01	1.96
76	3.97	3.12	2.72	2.49	2.33	2.22	2.13	2.06	2.01	1.96
77	3.97	3.12	2.72	2.49	2.33	2.22	2.13	2.06	2.00	1.96
78	3.96	3.11	2.72	2.49	2.33	2.22	2.13	2.06	2.00	1.95
79	3.96	3.11	2.72	2.49	2.33	2.22	2.13	2.06	2.00	1.95
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
81	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	2.00	1.95
82	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	2.00	1.95
83	3.96	3.11	2.71	2.48	2.32	2.21	2.12	2.05	1.99	1.95
84	3.95	3.11	2.71	2.48	2.32	2.21	2.12	2.05	1.99	1.95
85	3.95	3.10	2.71	2.48	2.32	2.21	2.12	2.05	1.99	1.94
86	3.95	3.10	2.71	2.48	2.32	2.21	2.12	2.05	1.99	1.94
87	3.95	3.10	2.71	2.48	2.32	2.20	2.12	2.05	1.99	1.94
88	3.95	3.10	2.71	2.48	2.32	2.20	2.12	2.05	1.99	1.94
89	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
91	3.95	3.10	2.70	2.47	2.31	2.20	2.11	2.04	1.98	1.94
92	3.94	3.10	2.70	2.47	2.31	2.20	2.11	2.04	1.98	1.94
93	3.94	3.09	2.70	2.47	2.31	2.20	2.11	2.04	1.98	1.93
94	3.94	3.09	2.70	2.47	2.31	2.20	2.11	2.04	1.98	1.93
95	3.94	3.09	2.70	2.47	2.31	2.20	2.11	2.04	1.98	1.93
96	3.94	3.09	2.70	2.47	2.31	2.19	2.11	2.04	1.98	1.93
97	3.94	3.09	2.70	2.47	2.31	2.19	2.11	2.04	1.98	1.93
98	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.98	1.93
99	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.98	1.93
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
Inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Table E: 10% Critical Values for the F distribution

df2/df1	1	2	3	4	5	6	7	8	9	10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
31	2.87	2.48	2.27	2.14	2.04	1.97	1.92	1.88	1.84	1.81
32	2.87	2.48	2.26	2.13	2.04	1.97	1.91	1.87	1.83	1.81
33	2.86	2.47	2.26	2.12	2.03	1.96	1.91	1.86	1.83	1.80
34	2.86	2.47	2.25	2.12	2.02	1.96	1.90	1.86	1.82	1.79
35	2.85	2.46	2.25	2.11	2.02	1.95	1.90	1.85	1.82	1.79
36	2.85	2.46	2.24	2.11	2.01	1.94	1.89	1.85	1.81	1.78
37	2.85	2.45	2.24	2.10	2.01	1.94	1.89	1.84	1.81	1.78
38	2.84	2.45	2.23	2.10	2.01	1.94	1.88	1.84	1.80	1.77
39	2.84	2.44	2.23	2.09	2.00	1.93	1.88	1.83	1.80	1.77
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
41	2.83	2.44	2.22	2.09	1.99	1.92	1.87	1.82	1.79	1.76
42	2.83	2.43	2.22	2.08	1.99	1.92	1.86	1.82	1.78	1.75
43	2.83	2.43	2.22	2.08	1.99	1.92	1.86	1.82	1.78	1.75
44	2.82	2.43	2.21	2.08	1.98	1.91	1.86	1.81	1.78	1.75
45	2.82	2.42	2.21	2.07	1.98	1.91	1.85	1.81	1.77	1.74
46	2.82	2.42	2.21	2.07	1.98	1.91	1.85	1.81	1.77	1.74
47	2.82	2.42	2.20	2.07	1.97	1.90	1.85	1.80	1.77	1.74
48	2.81	2.42	2.20	2.07	1.97	1.90	1.85	1.80	1.77	1.73
49	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73

df2/df1	1	2	3	4	5	6	7	8	9	10
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73
51	2.81	2.41	2.19	2.06	1.96	1.89	1.84	1.79	1.76	1.73
52	2.80	2.41	2.19	2.06	1.96	1.89	1.84	1.79	1.75	1.72
53	2.80	2.41	2.19	2.05	1.96	1.89	1.83	1.79	1.75	1.72
54	2.80	2.40	2.19	2.05	1.96	1.89	1.83	1.79	1.75	1.72
55	2.80	2.40	2.19	2.05	1.95	1.88	1.83	1.78	1.75	1.72
56	2.80	2.40	2.18	2.05	1.95	1.88	1.83	1.78	1.75	1.71
57	2.80	2.40	2.18	2.05	1.95	1.88	1.82	1.78	1.74	1.71
58	2.79	2.40	2.18	2.04	1.95	1.88	1.82	1.78	1.74	1.71
59	2.79	2.39	2.18	2.04	1.95	1.88	1.82	1.78	1.74	1.71
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
61	2.79	2.39	2.18	2.04	1.94	1.87	1.82	1.77	1.74	1.71
62	2.79	2.39	2.17	2.04	1.94	1.87	1.82	1.77	1.73	1.70
63	2.79	2.39	2.17	2.04	1.94	1.87	1.81	1.77	1.73	1.70
64	2.79	2.39	2.17	2.03	1.94	1.87	1.81	1.77	1.73	1.70
65	2.78	2.39	2.17	2.03	1.94	1.87	1.81	1.77	1.73	1.70
66	2.78	2.38	2.17	2.03	1.94	1.87	1.81	1.77	1.73	1.70
67	2.78	2.38	2.17	2.03	1.94	1.86	1.81	1.76	1.73	1.70
68	2.78	2.38	2.17	2.03	1.93	1.86	1.81	1.76	1.73	1.69
69	2.78	2.38	2.16	2.03	1.93	1.86	1.81	1.76	1.72	1.69
70	2.78	2.38	2.16	2.03	1.93	1.86	1.80	1.76	1.72	1.69
71	2.78	2.38	2.16	2.03	1.93	1.86	1.80	1.76	1.72	1.69
72	2.78	2.38	2.16	2.02	1.93	1.86	1.80	1.76	1.72	1.69
73	2.78	2.38	2.16	2.02	1.93	1.86	1.80	1.76	1.72	1.69
74	2.77	2.38	2.16	2.02	1.93	1.86	1.80	1.75	1.72	1.69
75	2.77	2.37	2.16	2.02	1.93	1.85	1.80	1.75	1.72	1.69
76	2.77	2.37	2.16	2.02	1.92	1.85	1.80	1.75	1.72	1.68
77	2.77	2.37	2.16	2.02	1.92	1.85	1.80	1.75	1.71	1.68
78	2.77	2.37	2.16	2.02	1.92	1.85	1.80	1.75	1.71	1.68
79	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68
80	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68
81	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68
82	2.77	2.37	2.15	2.01	1.92	1.85	1.79	1.75	1.71	1.68
83	2.77	2.37	2.15	2.01	1.92	1.85	1.79	1.75	1.71	1.68
84	2.77	2.37	2.15	2.01	1.92	1.85	1.79	1.74	1.71	1.68
85	2.77	2.37	2.15	2.01	1.92	1.84	1.79	1.74	1.71	1.67
86	2.76	2.37	2.15	2.01	1.92	1.84	1.79	1.74	1.71	1.67
87	2.76	2.36	2.15	2.01	1.91	1.84	1.79	1.74	1.70	1.67
88	2.76	2.36	2.15	2.01	1.91	1.84	1.79	1.74	1.70	1.67
89	2.76	2.36	2.15	2.01	1.91	1.84	1.79	1.74	1.70	1.67
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
91	2.76	2.36	2.14	2.01	1.91	1.84	1.78	1.74	1.70	1.67
92	2.76	2.36	2.14	2.01	1.91	1.84	1.78	1.74	1.70	1.67
93	2.76	2.36	2.14	2.01	1.91	1.84	1.78	1.74	1.70	1.67

df2/df1	1	2	3	4	5	6	7	8	9	10
94	2.76	2.36	2.14	2.01	1.91	1.84	1.78	1.74	1.70	1.67
95	2.76	2.36	2.14	2.00	1.91	1.84	1.78	1.74	1.70	1.67
96	2.76	2.36	2.14	2.00	1.91	1.84	1.78	1.74	1.70	1.67
97	2.76	2.36	2.14	2.00	1.91	1.84	1.78	1.73	1.70	1.67
98	2.76	2.36	2.14	2.00	1.91	1.84	1.78	1.73	1.70	1.66
99	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.70	1.66
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66
Inf	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

Table F: 5% One-sided Critical Values for the Durbin-Watson Distribution

N	K = 1		K = 2		K = 3		K = 4		K = 5		K = 6		K = 7	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.81	1.75	0.69	1.97	0.56	2.21	0.45	2.47	0.34	2.73
16	1.11	1.37	0.98	1.54	0.86	1.73	0.73	1.93	0.62	2.15	0.50	2.39	0.40	2.62
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.66	2.10	0.55	2.32	0.45	2.54
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06	0.60	2.26	0.50	2.46
19	1.18	1.40	1.07	1.53	0.97	1.68	0.86	1.85	0.75	2.02	0.65	2.21	0.55	2.40
20	1.20	1.41	1.10	1.54	1.00	1.68	0.89	1.83	0.79	1.99	0.69	2.16	0.60	2.34
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96	0.73	2.12	0.64	2.29
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94	0.77	2.09	0.68	2.25
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92	0.80	2.06	0.72	2.21
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90	0.84	2.04	0.75	2.17
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89	0.87	2.01	0.78	2.14
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88	0.90	1.99	0.82	2.12
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.00	1.86	0.93	1.97	0.85	2.09
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85	0.95	1.96	0.87	2.07
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84	0.98	1.94	0.90	2.05
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83	1.00	1.93	0.93	2.03
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83	1.02	1.92	0.95	2.02
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82	1.04	1.91	0.97	2.00
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81	1.06	1.90	0.99	1.99
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.14	1.81	1.08	1.89	1.02	1.98
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80	1.10	1.88	1.03	1.97
36	1.41	1.52	1.35	1.59	1.30	1.65	1.24	1.73	1.18	1.80	1.11	1.88	1.05	1.96
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80	1.13	1.87	1.07	1.95
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.20	1.79	1.15	1.86	1.09	1.94
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79	1.16	1.86	1.10	1.93
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79	1.18	1.85	1.12	1.93
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78	1.24	1.84	1.19	1.90
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77	1.29	1.82	1.25	1.88
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.37	1.77	1.33	1.81	1.29	1.86
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77	1.37	1.81	1.34	1.85
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77	1.40	1.81	1.37	1.84
70	1.58	1.64	1.55	1.67	1.53	1.70	1.49	1.74	1.46	1.77	1.43	1.80	1.40	1.84
75	1.60	1.65	1.57	1.68	1.54	1.71	1.52	1.74	1.49	1.77	1.46	1.80	1.43	1.83
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77	1.48	1.80	1.45	1.83
85	1.62	1.67	1.60	1.70	1.58	1.72	1.55	1.75	1.53	1.77	1.50	1.80	1.47	1.83
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78	1.52	1.80	1.49	1.83
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78	1.54	1.80	1.51	1.83
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78	1.55	1.80	1.53	1.83

Source: N. E. Savin and Kenneth J. White, "The Durbin–Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors," *Econometrica*, November 1977, p. 1994. Reprinted with permission.

Note: N = number of observations, K = number of explanatory variables excluding the constant term. We assume that the equation contains a constant term and no lagged dependent variables.