Diego Escobari

The University of Texas - Pan American

Introduction to Econometrics

(preliminary class notes)

ECON 3341

February 16, 2012

Contents

Chapter 1
Random Variables, Sampling and Estimation

1.1 Introduction

This chapter will cover the most important basic statistical theory you need in order to understand the econometric material that will be coming in the next chapters. The key topics that we will review are the following:

- *Descriptive statistics*. e.g. mean and variance.
- *Probability*. e.g. events, relative frequency, marginal and conditional probability distributions.
- *Random variables, probability distributions, and expectations*.
- *Sampling*. e.g. simple random sampling.
- *Estimation*. e.g. the distinction between and estimator and an estimate.
- *Statistical inference*. t and F tests.

1.2 Probabilities

1.2.1 Events

Random experiment. Process leading to two or more possible outcomes, with uncertainty as to which outcome will occur.

Flip of a coin, toss of a die, a students takes a class and either obtains an A or not.

Sample space. Set of all basic outcomes of a random experiment.

When flipping a coin, S = [head, tail].
When taking a class, S = [A, B, C, D, F, drop].
When tossing a die, S = [1, 2, 3, 4, 5, 6].
No two outcomes can occur simultaneously.

Event. Subset of basic outcomes in the sample space.

Event E_1: "Pass the class" then the subset of basic outcomes is A, B, C.

Intersection of event. When two events E_1 and E_2 have some basic outcomes in common. It is denoted by $E_1 \cap E_2$.

Event E_1: Individuals with college degree.
Event E_2: Individuals who are married.
$E_1 \cap E_2$: Individuals who have college degree and are married.

Joint probability. Probability that the intersection occurs.

Mutually exclusive events. E_1 and E_2 are mutually exclusive if $E_1 \cap E_2$ is empty.

Union of events. Denoted by $E_1 \cup E_2$. At least one of these events occurs. Either E_1, E_2, or both.

Complement. The complement of E is denoted by $\bar{E}$ and it is the set of basic outcomes of a random experiment that belongs to S, but not to E_1.

E_1 is the complement of $\bar{E}_1$
Event E_2: Individuals who are married.
E_1 and $\bar{E}$ are mutually exclusive events.

1.2.2 *Probability postulates*

Given a random experiment, we want to determine the probability that a particular event will occur. A probability is a measure from 0 to 1.

$0 \rightarrow$ the event will not occur.
$1 \rightarrow$ the event is certain.

When the outcomes are equally likely to occur, the probability of an event E is:

$P(E) = N_E/N$
N_E: Number of outcomes in event E.
N: Total number of outcomes in the sample space S.

Example 1: Flip of a coin, Event E is "head" then P(E) = 1/2. $N_E = 1$ and N = 2.

Example 2: Event E is "winning the lottery" then if there are 1000 lottery tickets and you bought, 2 P(E) = 2/1000 = 0.002.

Some probability rules

$P(E \cup \bar{E}) = P(E) + P(\bar{E}) = 1$.
$P(\bar{E}) = 1 - P(E)$.

Conditional probability

$P(E_1 \mid E_2)$: Probability that E_1 occurs, given that E_2 has already occurred.
$P(E_1 \mid E_2) = P(E_1 \cap E_2) / P(E_2)$ given that $P(E_2) > 0$.

Addition rule

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

Statistically independent events

$P(E_1 \cap E_2) = P(E_1)P(E_2)$.
$P(E_1 \mid E_2) = P(E_1)P(E_2) / P(E_2) = P(E_1)$.

1.3 Discrete random variables and expectations

1.3.1 Discrete random variables

Random variable. Variable that takes numerical values determined by the outcome of a random experiment.

Examples: Hourly wage, GDP, inflation, the number when tossing a die.
Notation: Random variable X can take n possible values $x_1, x_2, \cdots x_n$.

Discrete random variable. A random variable that takes a countable number of values.

Examples: Number of years of education.

Continuous random variable. A random variable that can take any value on an internal.

Examples: Wage, GDP, exact weight.

Consider tossing two dies (green and red). This will yield 36 possible outcomes because the green can take 6 possible values and the red can take also 6 values, $6 \times 6 = 36$. Let's define the random variable X to be the sum of two dice. Therefore X can take 11 possible values, from 2 to 12. This information is summarized in the following tables.

Table 1.1 Outcomes with two dies

red / green	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Table 1.2 Frequencies and probability distributions

Value of X	2	3	4	5	6	7	8	9	10	11	12
Frequency	1	2	3	4	5	6	5	4	3	2	1
Probability (p)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

1.3.2 *Expected value of random variables*

Let $E(X)$ be the expected value of the random variable X. The expected value of a discrete random variable is the weighted average of all its possible values, taking the probability of each outcome as its weight. Random variable X can take n particular values $x_1, x_2, \ldots, x_n$ and the probability of x_i is given by p_i. Then we have that the expected value is given by:

$$E(X) = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n = \sum_{i=1}^{n} x_i p_i. \tag{1.1}$$

We can also write the expected value as: $E(X) = \mu_X$. For the previous example we can calculate that the expected value as:

$$E(X) = 2 \cdot 1/36 + 3 \cdot 2/36 + \cdots + 12 \cdot 1/36 = 252/36 = 7 \tag{1.2}$$

Table 1.3 Expected value of X, two dice example

X	p	Xp
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36
Total	$E(X) = \sum_{i=1}^{n} x_i p_x$	252/36 = 7

1.3.3 Expected value rules

Let X, Y, and Z denote three random variables, and let b, b_1, and b_2 be arbitrary constants. Then,

$$E(X+Y+Z) = E(X) + E(Y) + E(Z) \tag{1.3}$$

$$E(bX) = bE(X) \quad \text{for a constant} \quad b \tag{1.4}$$

$$E(b) = b \tag{1.5}$$

For the example where $Y = b_1 + b_2 X$, b_1 and b_2 are constants we want to calculate $E(X)$.

$$\begin{aligned} E(Y) &= E(b_1 + b_2 X) \\ &= E(b_1) + E(b_2 X) \\ &= b_1 + b_2 E(X) \end{aligned} \tag{1.6}$$

1.3.4 Variance of a discrete random variable

Let $\text{var}(X)$ be the variance of the random variable X. $\text{var}(X)$ is a useful measure of the dispersion of its probability distribution. It is defined as the expected value of the square of the difference between X and its mean. That is, $E[(X - \mu_X)^2]$, where

μ_X is the population mean of X.

$$\begin{aligned}\mathrm{var}(X) &= \sigma_X^2 = E[(X-\mu_X)^2] & (1.7)\\ &= (x_1-\mu_X)^2 p_1 + (x_2-\mu_X)^2 p_2 + \cdots + (x_n-\mu_X)^2 p_n & (1.8)\\ &= \sum_{i=1}^{n}(x_i-\mu_X)^2 p_i\end{aligned}$$

Taking the square root of the variance (σ_X^2) one can obtain the standard deviation, σ_X. The standard deviation also serves as a measure of dispersion of the probability distribution. A useful way to write the variance is:

$$\sigma_X^2 = E(X^2) - \mu_X^2. \tag{1.9}$$

From the previous example of tossing two dies, we have that the population variance can be calculated as follows:

Table 1.4 Population variance, X from the two dice example

X	p	$X-\mu_X$	$(X-\mu_X)^2$	$(X-\mu_X)^2 p$
2	1/36	-5	25	0.69
3	2/36	-4	16	0.89
4	3/36	-3	9	0.75
5	4/36	-2	4	0.44
6	5/36	-1	1	0.14
7	6/36	0	0	0.00
8	5/36	1	1	0.14
9	4/36	2	4	0.44
10	3/36	3	9	0.75
11	2/36	4	16	0.89
12	1/36	5	25	0.69
Total				5.83

1.3.5 Probability density

Because discrete random variables, by definition, can only take a finite number of values, they are easy to summarize graphically. The probability distribution is the graph that links all the values that a random variable can take with its corresponding probabilities. For the two dice example above, see Figure 1.1.

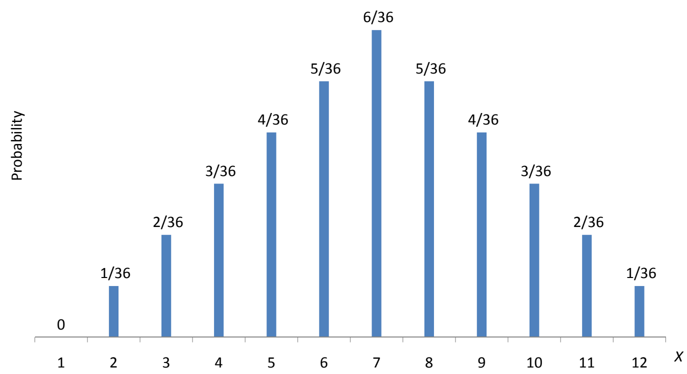

Fig. 1.1 Discrete probabilities, X from the two dice example

1.4 Continuous random variables

1.4.1 Probability density

Continuous random variables can take any value on an interval. This means that it can take an infinite number of different values, hence it is not possible to obtain a graph like the one presented in Figure 1.1 for a continuous random variable. Instead, we will define the probability of a random variable lying within a given interval. For example, the probability that the height of an individual is between 5.5 and 6 feet. This is depicted in Figure 1.2 as the shaded area below the probability density curve for the values of X between 5.5 and 6. The probability of the random variable X written as a function of the random variable is known as the probability density function. We can write this ones as $f(X)$. Then, if we use a little math we can easily find the area under the curve. Recall that the are under a curve can be obtained by taking the integral.

Probability density function. Is a function that describes the relative likelihood for a random variable to occur at a given point.

$$\int_{5.5}^{6} f(X) = 0.18 \qquad (1.10)$$
$$\int_{0}^{\infty} f(X) = 1$$

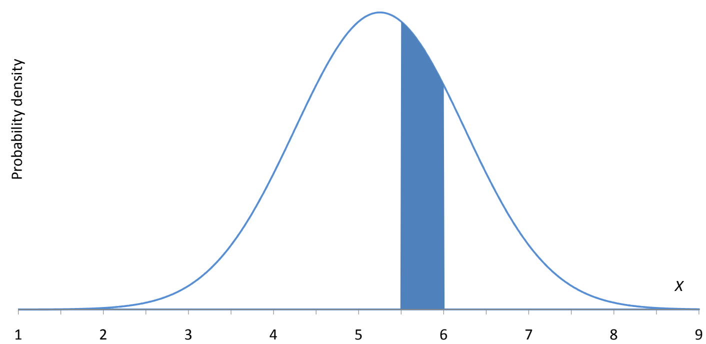

Fig. 1.2 Continuous probabilities, X from the height example

The first line in the equation above just calculates the integral under the curve $f(X)$ between the points 5.5 and 6. The second line shows that the whole area under the curve presented in Figure 1.2 is equal to one. This is for the same reason why the summation of all the bars in Figure 1.1 are also equal to one; the total probability is always equal to one.

1.4.2 Normal distribution

The normal distribution is the most widely known continuous probability distribution. The graph associated with its probability density function has a bell-shape and its is known as the Gaussian function or bell curve. Its probability density function is given by:

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu^2)}{2\sigma^2}} \tag{1.11}$$

where μ is the mean and σ^2 is the variance. Figure 1.2 is an example of this distribution.

1.4.3 *Expected value and variance of a continuous random variable*

The basic difference between a discrete and a continuous random variable is that the second can take on infinite possible values, hence the summations signs that are used to calculate the expected value and the variance of a discrete random variable cannot be used for a continuous random variable. Instead, we use integral signs. For the expected value we have:

$$E(X) = \int X f(X) dX \tag{1.12}$$

where the integration is performed over the interval for which $f(X)$ is defined. For the variance we have:

$$\sigma_X^2 = E[(X - \mu_X)^2] = \int (X - \mu_X)^2 f(X) dX \tag{1.13}$$

1.5 Covariance and correlation

1.5.1 *Covariance*

When dealing with two variables, the first question you want to answer is whether these variables move together or whether they move in opposite directions. The covariance will help us answer that question. For two random variables X and Y, the covariance is defined as:

$$\text{cov}(X,Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \tag{1.14}$$

where μ_X and μ_Y are the population means of X and Y, respectively. When to random variables are independent, their covariance is equal to zero. When $\sigma_{XY} > 0$ we say that the variables move together. When $\sigma_{XY} < 0$ they move in opposite directions.

1.5.2 *Correlation*

One concern when using the $\text{cov}(X,Y)$ as a measure of association is that the result is measured in the units of X times the units of Y. The correlation coefficient, that is dimensionless, overcomes this difficulty. For variables X and Y the correlation coefficient is defined as:

$$\mathrm{corr}(X,Y) = \rho_{YX} = \frac{\sigma_{YX}}{\sqrt{\sigma_X^2 \sigma_Y^2}} \tag{1.15}$$

The correlation coefficient is a number between -1 and 1. When it is positive, we say that there is a positive correlation between X and Y and that these two variables move in the same direction. When it is negative, we say that they move in opposite directions.

1.6 Sampling and estimators

Notice that in the two dice example we know the population characteristics, that is, the probability distribution. From this probability distribution it is easy to obtain the population mean an variance. However, what happens most of the time is that we need to rely on a data set to get estimates of the population parameters (e.g the mean and the variance). In that case the estimates of the population parameters are obtained using estimators, and the sample needs to have certain characteristics. The estimators and the sampling are the subject of this section.

1.6.1 Sampling

The most common way to obtain a sample from the population is through simple random sampling.

Simple random sampling. It is a procedure to obtain a sample from the population, where each of the observations is chosen randomly and entirely by chance. This means that each observation in the population has the same probability of being chosen.

Once the sample of the random variable X has be generated, each of the n observations can be denoted by $\{x_1, x_2, \cdots, x_n\}$.[1]

[1] The textbook Dougherty (2007) makes the distinction between the specific values of the random variable X before and after they are known, and emphasizes this distinction by using uppercase and lowercase letter. This distinction is useful only in some cases and that is why most textbooks do not make this distinction. We will follow emphasize the distinction and we will use only lowercase letters.

1.6.2 Estimators

Estimator. It is a general rule (mathematical formula) for estimating an unknown population parameter given a sample of data.

For example, an estimator for the population mean is the sample mean:

$$\bar{X} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{1.16}$$

An interesting feature of this estimator is that the variance of $\bar{X}$ is $1/n$ times the variance of X. The derivation is the following:

$$\sigma_{\bar{X}}^2 = \mathrm{var}(\bar{X}) \tag{1.17}$$

$$\sigma_{\bar{X}}^2 = \mathrm{var}\{\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\} \tag{1.18}$$

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2}\mathrm{var}\{\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\} \tag{1.19}$$

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2}\{\mathrm{var}(x_1) + \mathrm{var}(x_2) + \cdots + \mathrm{var}(x_n)\} \tag{1.20}$$

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2}\{\sigma_X^2 + \sigma_X^2 + \cdots + \sigma_X^2\} \tag{1.21}$$

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2}\{n\sigma_X^2\} = \frac{\sigma_X^2}{n} \tag{1.22}$$

Graphically, this result is shown in Figure 1.3. The distribution of X has a higher variance (it is more dispersed) than the distribution of $\bar{X}$.

1.7 Unbiasedness and efficiency

1.7.1 Unbiasedness

Because estimators are random variables, we can take expectations of the estimators. If the expectation of the estimator is equal to the true population parameter, then we say that this estimator is unbiased. Let θ be the population parameter and let $\hat{\theta}$ be a point estimator of θ. Then, $\hat{\theta}$ is unbiased if:

$$E(\hat{\theta}) = \theta \tag{1.23}$$

Example. The sample mean of X is an unbiased estimator of the population mean μ_X:

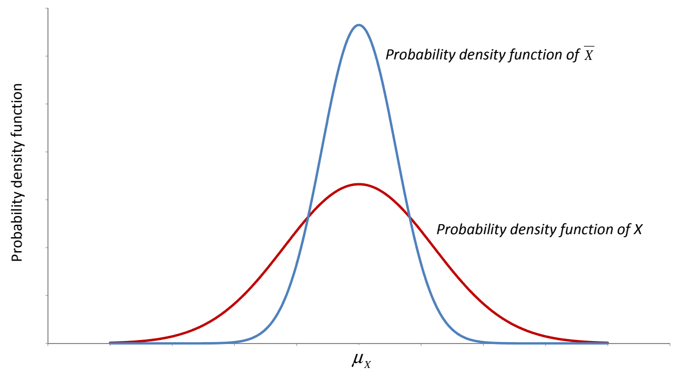

Fig. 1.3 Probability density functions of X and $\bar{X}$.

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n}E(\sum_{i=1}^{n} x_i) \tag{1.24}$$
$$= \frac{1}{n}\sum_{i=1}^{n}(E(x_i)) = \frac{1}{n}\sum_{i=1}^{n}\mu_X = \frac{1}{n}n\mu_X = \mu_X$$

Unbiased estimator. An estimator is unbiased if its expected value is equal to the true population parameter.

The bias of an estimator is just the difference between its expected value and the true population parameter:

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta \tag{1.25}$$

1.7.2 Efficiency

It is not only important that an estimator is on average correct (unbiased), but also that it has a high probability of being close to the true parameter. When comparing two estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, we say that $\hat{\theta}_1$ is more efficient if $\mathrm{var}(\hat{\theta}_1) < \mathrm{var}(\hat{\theta}_2)$. A comparison of the efficiency between these two estimators in presented in Figure 1.4. The estimator with higher variance, $(\hat{\theta}_2)$, is more dispersed.

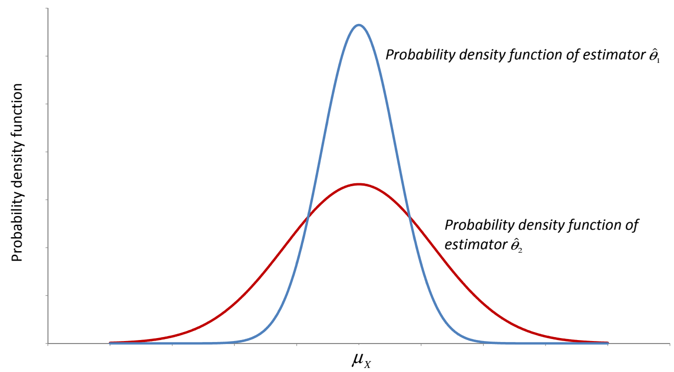

Fig. 1.4 Efficiency of estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, with $\text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$.

Most efficient estimator. The estimator with the smallest variance from all unbiased estimators.

1.7.3 Unbiasedness versus efficiency

Both, unbiasedness and efficiency, are desired properties of an estimator. However, there may be conflicts in the selection between two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, if, for example, $\hat{\theta}_1$ is more efficients, but it is also biased. This case is presented in Figure 1.5.

The simplest way to select between these two estimators is to pick the one that yields the smallest mean square error (MSE):

$$\text{MSE}(\hat{\theta}) = \text{var}(\hat{\theta}) + bias(\hat{\theta})^2 \tag{1.26}$$

1.8 Estimators for the variance, covariance, and correlation

While we have already seen the populations formulas for the variance, covariance and correlation, it is important to keep in mind that we do not have the whole population. The data sets we will be working with are just samples of the populations. The formula for the sample variance is:

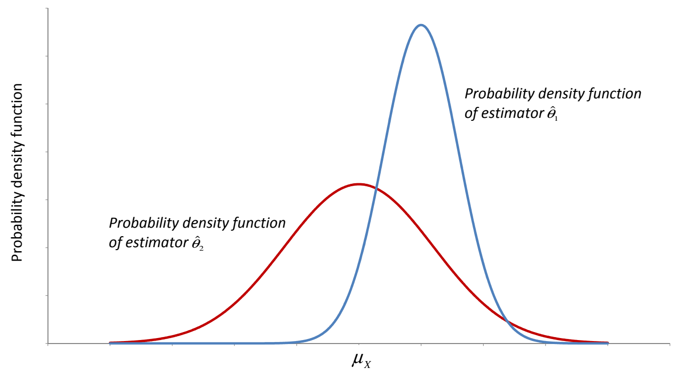

Fig. 1.5 $\hat{\theta}_2$ is unbiased, but $\hat{\theta}_1$ is more efficient.

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2 \tag{1.27}$$

Notice how we changed the notation from σ^2 to s^2. The first one denotes the population variance, while the second one refers to the sample variance. An estimator for the population covariance is given by:

$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y}). \tag{1.28}$$

Finally, the formula for the correlation coefficient, r_{XY}, is:

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2 \sum_{i=1}^{n}(y_i - \bar{Y})^2}}. \tag{1.29}$$

1.9 Asymptotic properties of estimators

Asymptotic properties of estimators just refers to their properties when the number of observations in the sample grows large and approached to infinity.

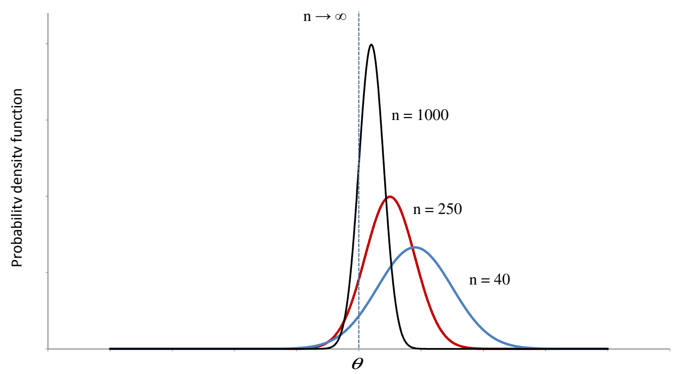

Fig. 1.6 The estimator is biased for small samples, but consistent.

1.9.1 Consistency

An estimator $\hat{\theta}$ is said to be consistent if its bias becomes smaller as the sample size grows large. Consistency is important because many of the most common estimators used in econometrics are biased, then the minimum we should expect from these estimators is that the bias becomes small as we are able to obtain larger data sets. Figure 1.6 illustrates the concept of consistency by showing how an estimator of the population parameter θ becomes unbiased as $n \to \infty$.

1.9.2 Central limit theorem

Having normally distributed random variables is important because we can then construct, for example, confidence intervals for its mean. However, what if a random variable does not follow a normal distribution? The central limit theorem gives us the answer.

Central limit theorem. States the conditions under which the mean of a sufficiently large number of independent random variables (with finite mean and variance) will be approximate a normal distribution.

Hence, even if we do not know the underlying distribution of a random variable, we will still be able to construct confidence intervals that will be approximately valid. In a numerical example, let's assume that the random variable X follows a

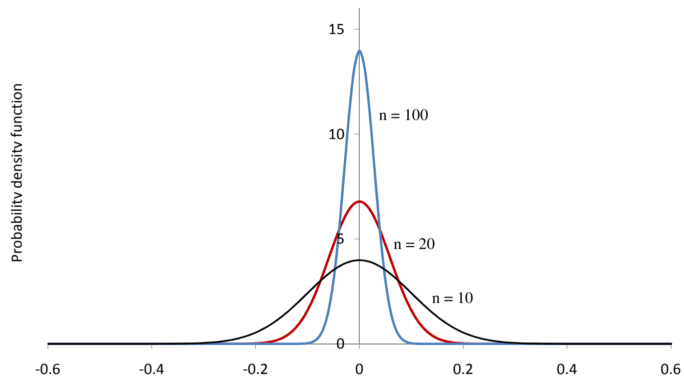

Fig. 1.7 Distribution of the sample mean of a uniform distribution.

uniform distribution [-0.5,0.5]. Hence, it is equally likely that this random variable takes any value within this range. Figure 1.7 shows the distribution of the average of this random variable for n = 10, 20, and 100. All of these three distributions look very close to a normal distribution.

Chapter 2
Simple Linear Regression

2.1 Simple linear model

The simple linear regression model shows how one known dependent variable is determined by a single explanatory variable (regressor). Is is written as:

$$Y_i = \beta_1 + \beta_2 X_i + u_i. \tag{2.1}$$

The subscript i refers to the observation $i = 1, 2, \ldots n$, and Y_i is the dependent variable. We break down Y_i into two components, the deterministic (nonrandom) component $\beta_1 + \beta_2 X_i$ and the stochastic (random) component u_i. The explanatory variable is X_i and the population parameters we want to estimate are given by intercept β_1 and the slope β_2. The term u_i is the disturbance term. Figure 2.1 shows a graphical representation of the problem. The regression line $Y_i = \beta_1 + \beta_2 X_i + u_i$ is shown as the upward sloping blue line. Only a single observation point at (X_i, Y_i) is illustrated. We can see how for this observation i, we break down Y_i into the disturbance term u_i given by the vertical distance between Y_i and $\hat{Y}_i$ and the height of the regression line at point X_i, given by $\beta_1 + \beta_2 X_i$.

2.2 Least squares regression

The main idea in econometric analysis is to estimate the parameters β_1 and β_2. The most popular estimator for these population parameters is the Ordinary Least Squares (OLS) estimator. Let the OLS estimators of β_1 and β_2 be b_1 and b_2, respectively. Then, the fitter regression equation is:

$$Y_i = b_1 + b_2 X_i + e_i. \tag{2.2}$$

The difference between Equations 2.1 and 2.2 is that the first correspond to the population, while the second is the sample counterpart. The idea in the OLS es-

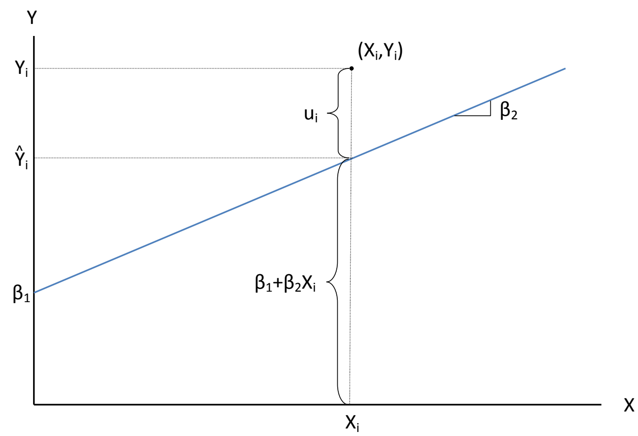

Fig. 2.1 Regression line $Y_i = \beta_1 + \beta_2 X_i + u_i$.

timator is simple, we want to pick values for the intercept b_1 and slope b_2 coefficients that are as close as possible to the actual data points. That is, we want to e_i ($e_i = Y_1 - b_1 - b_2 X_i$) to be small. Because some of the e_i are positive and some are negative, we will first square them to have all positive numbers. Then, to take into account all data points we will sum across all observations. That is how our objective is to pick b_1 and b_2 to minimize the following residual sum of squares:

$$RSS = e_1^2 + e_2^2 + \cdots + e_n^2 = \sum_{i=1}^{n} e_i^2. \tag{2.3}$$

This minimization exercise yields the OLS estimators:

$$b_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{2.4}$$

for the slope coefficient, and

$$b_1 = \bar{Y} - b_2 \bar{X} \tag{2.5}$$

for the intercept. The derivation of the least squares coefficient estimators (Equations 2.4 and 2.5) has the following steps. We start with the regression equation:

$$Y_i = b_1 + b_2 X_i + e_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$

For observation i we obtain the residual, then square it and finally sum across all observations to obtain the residual sum of squares:

$$\begin{aligned} e_i &= Y_i - \hat{Y}_i \\ e_i^2 &= (Y_i - \hat{Y}_i)^2 \\ \sum_{i=1}^{n} e_i^2 &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \end{aligned} \tag{2.6}$$

The coefficients b_1 and b_2 are chosen to minimize the residuals sum of squares:

$$\begin{aligned} &\min_{b_1,b_2} \quad \textstyle\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \\ &\min_{b_1,b_2} \textstyle\sum_{i=1}^{n}(Y_i - b_1 - b_2 X_i)^2 \end{aligned} \tag{2.7}$$

The first order necessary condition are:

$$-2\sum_{i=1}^{n}(Y_i - b_1 - b_2 X_i) = 0 \qquad \text{w.r.t.} \qquad b_1 \tag{2.8}$$

$$-2\sum_{i=1}^{n}X_i(Y_i - b_1 - b_2 X_i) = 0 \qquad \text{w.r.t.} \qquad b_2 \tag{2.9}$$

Dividing Equation 2.9 by n and working through some math we obtain the OLS estimators for the constant:

$$b_1 = \bar{Y} - b_2\bar{X}.$$

Plugging this result into Equation 2.9 we obtain:

$$b_2 = \frac{\sum_{i=0}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=0}^{n}(X_i - \bar{X})^2}.$$

2.3 Interpretation of the regression coefficients

If the estimated regression equation is given by:

$$\widehat{wage}_i = 4.64 + 0.09 exper_i, \tag{2.10}$$

where *wage* is the hourly wage measured in dollars, and *exper* is the number of years of experience, then the interpretation of the slope coefficient is the following:

$$\frac{\Delta wage}{\Delta exper} = 0.09.$$

Therefore, if the change in the number of years of experience is one, $\Delta exper$, then the change in the hourly wage in dollars is given by $\Delta wage = 0.09$. In words, an additional year of experience will increase your hourly wage by 0.09 dollars (or 9 cents). For the interpretation of the intercept, just consider the case where someone has not experience, $exper = 0$. Then, this person's predicted wage will be 4.64 dollars.

If the estimated regression equation takes the form:

$$\widehat{\log wage}_i = 1.38 + 0.02 exper_i, \tag{2.11}$$

where the $\log wage$ is the natural logarithm of $wage$, then the interpretation is different. Here, if the number of years of experience increases by one, the wage increases by 2% ($0.02\times$ 100 percent). Finally, for the folowing estimated equation:

$$\widehat{\log wage}_i = 0.98 + 0.26 \log exper_i. \tag{2.12}$$

A one percent increase in $exper$ will increase $wage$ by 0.25 percent. The 0.26 is interpreted as an elasticity.

2.4 Goodness of fit

How good is the regression equation in explaining the variation in variable Y? First we need a way to measure the total variation in Y. Let's try the sum of squared deviations about the sample mean of Y. That is,

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{2.13}$$

Now, let's start with a simple equality:

$$Y_i - \bar{Y} = Y_i - \bar{Y}.$$

If we add and subtract $\hat{Y}_i$ on the right hand side of the above equality, we have

$$\begin{aligned} Y_i - \bar{Y} &= Y_i - \bar{Y} + \hat{Y}_i - \hat{Y}_i \\ Y_i - \bar{Y} &= (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \end{aligned}$$

Squaring both sides of the equation and then summing across all observations i we obtain:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{2.14}$$

$$TSS = ESS + RSS. \tag{2.15}$$

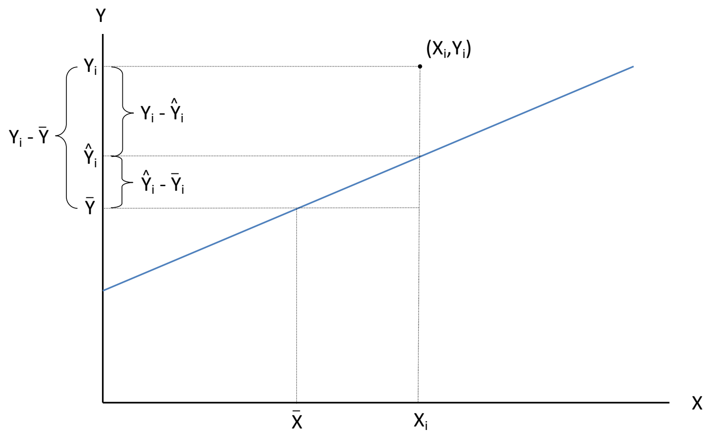

Fig. 2.2 Decomposition of $\hat{Y}_i - \bar{Y}$.

Notice that the sum of deviations from the mean is zero, that is why there are only two components on the right hand side. The TSS is the Total Sum of Squares, as presented in Equation 2.13. The first term on the right hand side is ESS, the Explained Sum of Squares, and the second term on the right hand side is the RRS, Residual Sum of Squares. This decomposition of the variable Y into two components can be appreciated in Figure 2.2. For every observation Y_i in the sample, the distance between Y_i and $\bar{Y}$ can be decomposed in two, the part that the regression equation can explain, $\hat{Y}_i - \bar{Y}$, and the part that the regression equation cannot explain, $Y_i - \hat{Y}_i$.

What is the proportion of the variation in Y that is explain by the regression equation? We just need to divide Equation 2.15 by TSS and define the ratio of ESS to TSS as the proportion of the explained variation in Y, the R^2:

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS} \tag{2.16}$$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \tag{2.17}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{2.18}$$

The R^2 is a number between zero and one, being higher when the model explains more of the variation in Y. Figures 2.3 and 2.4 illustrate how the regression line explain the variation in Y when the R^2 is low and high, respectively.

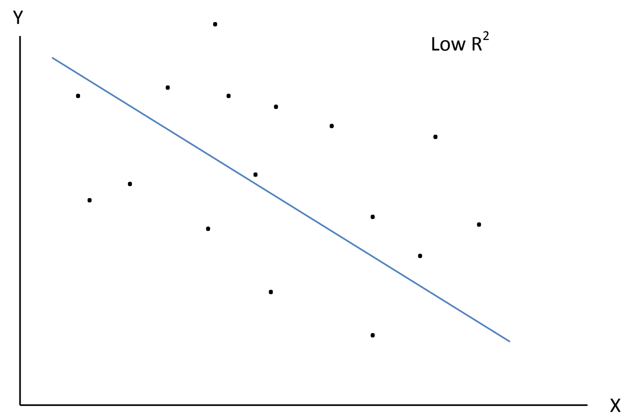

Fig. 2.3 Low R^2.

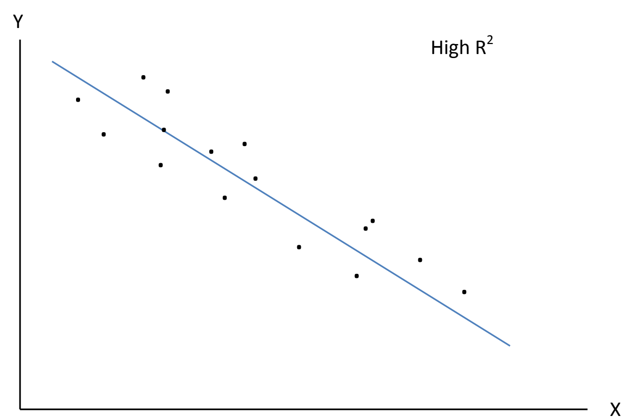

Fig. 2.4 High R^2.

Chapter 3
Properties and Hypothesis Testing

3.1 Types of data

The regression techniques developed in previous chapters can be applied to three different kinds of data.

1. Cross-sectional data.
2. Time series data.
3. Panel data.

The first consists on observing various economic unit (e.g. firms, countries, households, individuals) at one point in time. For example, we observe the wages, experience and education of many individuals, only once and at all at the same time. The second consists on observing the same economic unit at different point in time. For example, we observe daily stock prices over many years. Finally, the third combines the characteristics of the first and the second. That is, we observe various economic units at repeated points in time. For example, we have information about the inflation, unemployment and GDP of a group of countries and over many years.

3.2 Assumptions of the model

When the regressors in our econometric model are non stochastic, we will make the following six assumptions.

1. The model is linear in the parameters and it is correctly specified.

$$Y = \beta_1 + \beta_2 X + u \tag{3.1}$$

$$Y = \beta_1 X^{\beta_2} + u \tag{3.2}$$

Equation 2.1 is linear in β, while Equation 2.2 is not.

2. There is some variation in the regressor in the sample. We need variation in the variable X to identify the relationship. Consider the OLS estimator for β_2:

$$b_2 = \frac{\sum_{i=0}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=0}^{n}(X_i - \bar{X})^2}. \tag{3.3}$$

   If there is no variation in X, then the denominator is zero and we cannot obtain b_2.
3. The expected value of the disturbance term is zero.

$$E(u_i) = 0 \quad \text{for all} \quad i. \tag{3.4}$$

   Some u_i will be negative, some will be positive, but on average they will be zero. If a constant is included in the model, the condition is satisfied automatically.
4. The disturbance term is homoscedastic.
   Homoscedasticity means that the variance of the error terms u_i is constant across all observations i. Hence, we can write:

$$\sigma_{u_i}^2 = \sigma_u^2 \quad \text{for all} \quad i. \tag{3.5}$$

   Because the error term has zero mean (from assumption 3), then the population variance of u_i is equal to:

$$E(u_i^2) = \sigma_u^2 \quad \text{for all} \quad i. \tag{3.6}$$

   σ_u^2 is a population parameter, therefore it is unknown and need to be estimated.
5. The values of the disturbance terms have independent distributions.

$$u_i \text{ is distributed independently of } u_j \text{ for all } j \neq i. \tag{3.7}$$

   This means that there is no *autocorrrelation* in the error term. This means that the population covariance between u_i and u_j is zero:

$$\sigma_{u_i u_j} = 0. \tag{3.8}$$

   With assumptions 1 through 5, we says that OLS coefficients are BLUE: Best Linear Unbiased Estimators. They are best, because they have the smallest variance across all unbiased estimators.
6. The disturbance term has a normal distribution.

$$u_i \sim N[0, \sigma_u^2] \quad \text{for all} \quad i. \tag{3.9}$$

   The error term is distributed normal with mean zero and variance σ_u^2. This assumption becomes useful at the time of performing t tests, F tests, and constructing confidence intervals for β_1 and β_2 using the regression results. The justification for this assumption depends on the *central limit theorem*. This one state that if a random variable is the composite result of the effects of a large number of other random variables (that are not necessarily normal), it will have an approximately normal distribution.

3.3 Unbiasedness of the coefficients

Recall that an estimator $\hat{\theta}$ is unbiased if $E(\hat{\theta}) = \theta$. The expected value of the estimator is equal to the true population parameter. For the slope coefficient in the OLS regression we have:

$$\begin{aligned} b_2 &= \frac{\sum_{i=0}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=0}^{n}(X_i - \bar{X})^2} \\ &= \beta_2 + \frac{\sum_{i=0}^{n}(X_i - \bar{X})u_i}{\sum_{i=0}^{n}(X_i - \bar{X})^2} \\ &= \beta_2 + \sum_{i=1}^{n} a_i u_i \end{aligned} \tag{3.10}$$

where

$$a_i = \frac{(X_i - \bar{X})}{\sum_{i=0}^{n}(X_i - \bar{X})^2}. \tag{3.11}$$

Thus, this shows that b_2 is equal to its true value, β_2, plus a linear combination of the values of the error terms. If we take expectations of b_2 we have:

$$E(b_2) = E(\beta_2) + E\big(\sum_{i=1}^{n} a_i u_i\big) = \beta_2 + \sum_{i=1}^{n} E(a_i u_i) = \beta_2 + \sum_{i=1}^{n} a_i E(u_i) = \beta. \tag{3.12}$$

The term a_i goes out of the expectation because a_i is only a function of constant Xs. In addition, the last equality holds because $E(u_i) = 0$. Hence, b_2 is an unbiased estimator of β_2, $E(b_2) = \beta_2$.

3.4 Precision of the coefficients

We are also interested on how precise b_1 and b_2 are in estimating the population parameters β_1 and β_2. A measure of this precision are their population variances, given by:

$$\sigma_{b_1}^2 = \sigma_u^2\Big(\frac{1}{n} + \frac{\bar{X}}{\sum_{i=0}^{n}(X_i - \bar{X})^2}\Big), \quad \text{and} \tag{3.13}$$

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum_{i=0}^{n}(X_i - \bar{X})^2} \tag{3.14}$$

One concern in the implementation of the above formulas is that σ_u^2 is an unknown population parameter and need to be estimated. A natural estimator for this regression variance is the variance of the regression errors. Because the population regression errors u_i are also unknown, we use the sample counterparts e_i and adjust for the corresponding degrees of freedom. Hence, we have:

$$S_u^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2. \tag{3.15}$$

This S_u^2 is the unbiased estimator of σ_u^2, and $n-2$ are the degrees of freedom. We subtract two from the sample size because we are estimating two parameters: the regression constant and one slope coefficient. Then, we use the following formulas to estimate the standard errors of b_1 and b_2:

$$S_{b_1} = \sqrt{S_u^2\Big(\frac{1}{n} + \frac{\bar{X}}{\sum_{i=0}^{n}(X_i-\bar{X})^2}\Big)}, \quad \text{and} \tag{3.16}$$

$$S_{b_2} = \sqrt{\frac{S_u^2}{\sum_{i=0}^{n}(X_i-\bar{X})^2}}. \tag{3.17}$$

3.5 The Gauss-Markov theorem

The *Gauss-Markov theorem* simply states that when assumptions 1 through 5 above are satisfied, the OLS estimators are Best Linear Unbiased Estimators (BLUE) of the regression parameters. Best refers to smallest variance.

3.6 Hypotheses testing

Hypothesis testing is simply a method of making decisions using data. It starts with the formulation of the null and the alternative hypotheses and then uses some test statistics to assess the truth of the null hypothesis.

3.6.1 Formulation of the null hypothesis

The formulation of the null hypothesis starts with a relationship in mind. For example, that the percentage rate of price inflation (p) depends on the percentage rate of wage inflation (w) following the linear equation:

$$p_i = \beta_1 + \beta_2 w_i + u_i \tag{3.18}$$

Then, you want to test the hypothesis that the price inflation is equal to the wage inflation. This is denoted by H_0 and it is know as the *null hypothesis*. In addition, we also define an alternative hypothesis, denoted by H_1 and represents the conclusion of the test if the null hypothesis is rejected. For our example the null and the alternative hypothesis are written as:

$$H_0 : \beta_2 = 1 \tag{3.19}$$

$$H_1 : \beta_2 \neq 1 \tag{3.20}$$

In general, the null and alternative hypotheses are:

$$H_0 : \beta_2 = \beta_2^0 \tag{3.21}$$

$$H_1 : \beta_2 \neq \beta_2^0. \tag{3.22}$$

3.6.2 *t*-tests

Recall that β_2 is unknown and that we have to use the estimate b_2. Then, the decision rule to reject the null hypothesis should compare the estimate b_2 with the hypothesized value β_2^0. Intuitively, if the values are far apart, then there is evidence against the null. This comparison should take into account the fact that b_2 is subject to some sampling variation (it is not the actual β_2). We will use the following statistic:

$$z = \frac{b_2 - \beta_2^0}{\sigma_{b_2}} \tag{3.23}$$

The numerator is just the distance between the regression estimate and the hypothesized value, with the denominator is the standard deviation of b_2, given by the square root of the expression in Equation 3.14. z is the number of standard deviations between b_2 and β_2. For a known σ_{b_2}, this one follows a normal distribution. However σ_{b_2} is unknown and we need to use the estimate of the standard error of b_2. This one is given by S_{b_2} and it is presented in Equation 3.17. Then we use the following t-statistic:

$$t = \frac{b_2 - \beta_2^0}{S_{b_2}} \tag{3.24}$$

To know if the deviations between b_2 and β_2^0 are significantly large, we compare this t-statistic with the critical values from the table t distribution with $n-2$ degrees of freedom. The null hypothesis is not rejected if the following condition is met:

$$-t_{n-2,\alpha/2} \leq \frac{b_2 - \beta_2^0}{S_{b_2}} \leq t_{n-2,\alpha/2} \tag{3.25}$$

Where $t_{n-2,\alpha/2}$ is just the notation of the critical value than comes from the t distribution with $n-2$ degrees of freedom and at significance level α. The *significance*

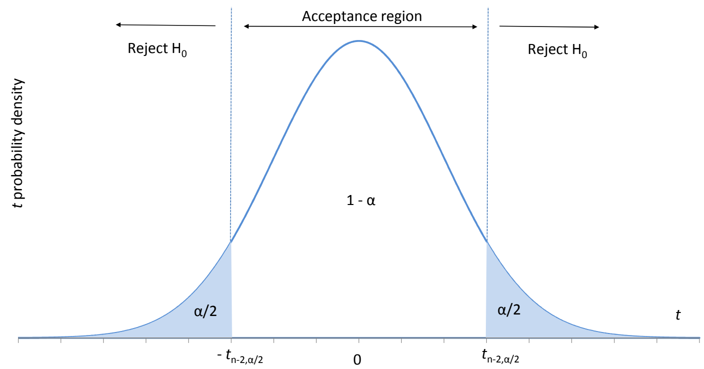

Fig. 3.1 Acceptance region for the t-test.

level is the probability that we reject the null hypothesis when in fact it is true. The rejection regions are illustrated in Figure 3.1.

3.6.3 Confidence intervals

The *confidence interval* indicates the reliability of an estimate. The confidence interval for the population parameter β_2 can be derived from Equation 3.25 in the following way:

$$1-\alpha = P\left(-t_{n-2,\alpha/2} \leq \frac{b_2-\beta_2}{S_{b_2}} \leq t_{n-2,\alpha/2}\right) \tag{3.26}$$

$$1-\alpha = P\left(-t_{n-2,\alpha/2} \cdot S_{b_2} \leq b_2-\beta_2 \leq t_{n-2,\alpha/2} \cdot S_{b_2}\right)$$

$$1-\alpha = P\left(b_2-t_{n-2,\alpha/2} \cdot S_{b_2} \leq \beta_2 \leq b_2+t_{n-2,\alpha/2} \cdot S_{b_2}\right)$$

The meaning of the above equation is that the population parameter β_2 will be between the lower confidence limit $b_2-t_{n-2,\alpha/2} \cdot S_{b_2}$ and the upper confidence limit $b_2+t_{n-2,\alpha/2} \cdot S_{b_2}$ with probability $(1-\alpha)$ or $100\times(1-\alpha)\%$. The *p values* provide an alternative approach to reporting the significance of regression coefficients or when carrying out more general hypothesis testing. As you can see from Equation 3.25 and Figure 3.1, different significance levels α can yield a different conclusion in the rejection or not of the null hypothesis. The p value of a hypothesis test represent the minimum significance level at which the null is rejected. Then, when the p value is below the significance level α we reject the null.

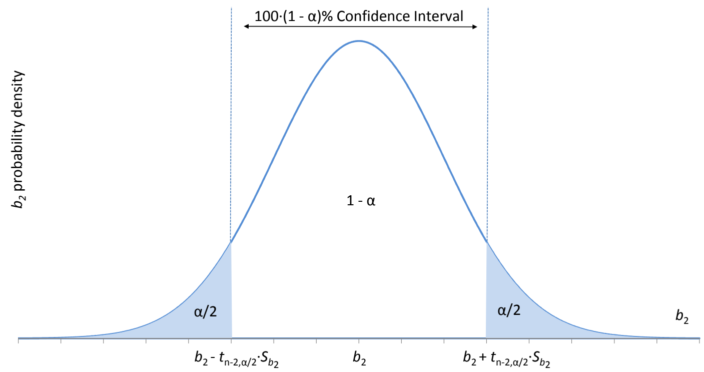

Fig. 3.2 Confidence interval for β_2.

3.6.4 F test

A useful tool if we want to test if there is no relationship between X and Y if the F test. In the simple linear regression model with only one slope coefficient, the null and the alternative in an F test are:

$$H_0 : \beta_2 = 0 \tag{3.27}$$

$$H_1 : \beta_2 \neq 0. \tag{3.28}$$

This test is build on the idea of testing how good is the regression model in explaining the variation in Y. In Equation 2.15 we already separated the variation of Y into its ‘explained’ and ‘unexplained’ components. These are:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{3.29}$$

$$TSS = ESS + RSS. \tag{3.30}$$

The total sum of squares (TSS) is the summation of the explained sum of squares (ESS) and the residual sum of squares (RSS). Then, the F *statistic* for goodness of fit of a regression is written as the explained sum of squares, per explanatory variable, divided by the residual sum of squares, per remaining degrees of freedom:

$$F = \frac{ESS/(k-1)}{RSS/(n-k)} \tag{3.31}$$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.2346
R Square	0.0551
Adjusted R Square	0.0543
Standard Error	4.5323
Observations	1260

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1505.5387	1505.5387	73.2906	0.0000
Residual	1258	25841.9006	20.5421		
Total	1259	27347.4393			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.6425	0.2326	19.9615	0.0000	4.1862	5.0988
X Variable 1	0.0914	0.0107	8.5610	0.0000	0.0705	0.1124

Fig. 3.3 Regression output in MS Excel.

where k is the total number of coefficients we are estimating, hence $(k-1)$ is the number of slope coefficients. That is, the total number of parameters we are estimating minus the constant parameter. If we divide the numerator and the denominator by TSS, then the F statistics can be written in terms of the R^2 as follows:

$$F = \frac{(ESS/TSS)/(k-1)}{(RSS/TSS)/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{3.32}$$

If this F statistic is greater that the critical value from the table F distribution with $(k-1)$ and $(n-k)$ degrees of freedom, $F_{k-1,n-k}$, we reject the null hypothesis and conclude that the regression model does not significantly explain the variation in variable Y. For the simple regression model with only one slope coefficient, $k=2$, we have:

$$F = \frac{R^2}{(1-R^2)/(n-2)}. \tag{3.33}$$

If this F statistic $> F_{1,n-2}$ we reject the null hypothesis presented in Equation 3.28.

3.7 Computer output

The computer regression output is very similar across different statistical packages. Figure 3.3 shows the output using MS Excel for the estimation of the following simple regression model:

$$wage_{=}\beta_1 + \beta_2 exper_i + u_i \tag{3.34}$$

To obtain the regression estimated coefficients we use Equations 2.4 and 2.5:

$$b_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = 0.091 \tag{3.35}$$

$$b_1 = \bar{Y} - b_2\bar{X} = 4.642 \tag{3.36}$$

The total sum of squares, estimates sum of squares, and residual sum of squares are obtained using 2.15 and 2.15:

$$TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 27347.439 \tag{3.37}$$

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = 1505.539 \tag{3.38}$$

$$RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = 25841.901 \tag{3.39}$$

The regression R^2 comes from Equation 2.18:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 0.055 \tag{3.40}$$

From the square root of Equation 3.15:

$$S_u = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} e_i^2} = 4.532 \tag{3.41}$$

Then, the standard errors of the coefficients are computer using Equations 3.17 and 3.17:

$$S_{b_1} = \sqrt{S_u^2\left(\frac{1}{n} + \frac{\bar{X}}{\sum_{i=0}^{n}(X_i - \bar{X})^2}\right)} = 0.233 \tag{3.42}$$

$$S_{b_2} = \sqrt{\frac{S_u^2}{\sum_{i=0}^{n}(X_i - \bar{X})^2}} = 0.011 \tag{3.43}$$

The F statistic uses Equation 3.32:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = 73.291 \tag{3.44}$$

The t statistics use Equation 3.24:

$$t = \frac{b_1}{S_{b_1}} = 19.961 \tag{3.45}$$

$$t = \frac{b_2}{S_{b_2}} = 8.561 \tag{3.46}$$

Finally, for the 95% upper and lower confidence levels, we use Equation 3.26:

$$b_1 - t_{n-2,\alpha/2} \cdot S_{b_1} = 4.186 \tag{3.47}$$

$$b_1 + t_{n-2,\alpha/2} \cdot S_{b_1} = 5.099 \tag{3.48}$$

$$b_2 - t_{n-2,\alpha/2} \cdot S_{b_2} = 0.071 \tag{3.49}$$

$$b_2 + t_{n-2,\alpha/2} \cdot S_{b_2} = 0.112 \tag{3.50}$$

Chapter 4
Multiple Regression Analysis

The simple linear regression covered in Chapter 2 can be generalized to include more than one variable. Multiple regression analysis is an extension of the simple regression analysis to cover cases in which the dependent variable is hypothesized to depend on more than one explanatory variable. While much of the analysis is an extension of the simple case, we have two main complications. (1) We need to discriminate between the effects of one variable and the effects of the other explanatory variables. (2) We have to decide which variables to include in the regression equation. In this chapter we will focus on the extension of the linear regression model and in (1). In a later chapter we will discuss (2).

4.1 Interpretation of the coefficients

Consider the following population multiple regression model with $(k-1)$ regressors:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + u. \tag{4.1}$$

A simple example of a multiple regression model is:

$$CRIME_i = \beta_1 + \beta_2 POPULATION_i + \beta_3 UNEMPLOY_i + \beta_4 POLICE_i + u_i, \tag{4.2}$$

where i refers to the city, $CRIME$ is crime rates, $POPULATION$ is just the number people in city i, $UNEMPLOY$ is the unemployment rate, and $POLICE$ is the number of police officers. To estimate the βs in Equation 4.2 you may need to observe crime rates and all the other variables for n cities. As before, u is the disturbance term. Because we have more that one regressor, the simple two dimensional characterization illustrated in Figure 2.1 is no longer applicable. Now, we have a $(k-1)$ dimensional problem. In our crime example we would need to have a 4D graph!

The sample counterpart of Equation 4.2 is:

$$CRIME_i = b_1 + b_2 POPULATION_i + b_3 UNEMPLOY_i + b_4 POLICE_i + e_i, \tag{4.3}$$

where the bs are the sample estimates of the βs, and are estimated using computer software via Ordinary Least Squares. We also express this relationship as the 4D fitted plane:

$$\widehat{CRIME}_i = b_1 + b_2 POPULATION_i + b_3 UNEMPLOY_i + b_4 POLICE_i. \tag{4.4}$$

Notice that we no longer write the disturbance term. Moreover, $\widehat{CRIME}_i$ is the fitted or predicted value of $CRIME_i$. The interpretation of the coefficients is the same as before. If the number of police officers increases by one, then the crime rate will change by b_4. Similar interpretation follows for b_2 and b_3.

4.2 Ordinary Least Squares

The OLS estimates are obtained in the same fashion as before. The unknown relationship is given by:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i. \tag{4.5}$$

The fitted OLS regression is:

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + \cdots + b_k X_{ki}. \tag{4.6}$$

Then, the OLS regression residuals are:

$$e_i = Y_i - \hat{Y} = Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i} - \cdots - b_k X_{ki}. \tag{4.7}$$

Recall that OLS minimizes the sum of squared residuals

$$\min_{b_1, b_2, \ldots, b_k} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2, \tag{4.8}$$

where $RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$ is the sum of squared residuals. We need to take the derivative of the RSS with respect to $b_1, b_2, \ldots, b_k$ and obtain k first order conditions. This yields a system of k equations with k unknowns, where the solution is the OLS estimators of the βs.

4.3 Assumptions

1. The model is linear in the parameters and correctly specified

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + u. \tag{4.9}$$

2. There is no exact linear relationship among the regressors in the sample. This is called multicollinearity.
3. The disturbance term has expectation zero

$$E(u_i) = 0 \qquad \text{for all} \quad i. \tag{4.10}$$

4. The disturbance term is homoscedastic.

$$\sigma_{u_i}^2 = \sigma_u^2 \qquad \text{for all} \quad i. \tag{4.11}$$

5. The values of the disturbance term have independent distributions.

$$u_i \text{ is distributed independently of } u_{i'} \text{ for all } i' \neq i. \tag{4.12}$$

6. The distribution term has a normal distribution.

$$u_i \sim N[0, \sigma^2] \qquad \text{for all} \quad i. \tag{4.13}$$

All the Xs are nonstochastic.

4.4 Properties of the coefficients

4.4.1 Unbiasedness

The OLS estimator b_j of β_j is unbiased:

$$E(b_j) = \beta_j \tag{4.14}$$

4.4.2 Efficiency

Following the results from the Gauss-Markov theorem, we have that OLS yields the most efficient linear estimators, in the sence that they are the one with the smallest variance among all linear estimators.

4.4.3 Precision of the coefficient, t tests, and confidence intervals

Beside our interest on the point estimates, we are also interested in performing hypotheses testing and building confidence intervals. To do this we need a measure of the precision of the coefficients. While we will not show the derivation here (as it required matrix algebra), each of the b_j has an standard error, S_{b_j}.

The null and alternative hypotheses about population coefficient j is written as:

$$H_0 : \beta_j = \beta_j^0 \tag{4.15}$$

$$H_1 : \beta_j \neq \beta_j^0. \tag{4.16}$$

which can be tested using the following t-statistic:

$$t = \frac{b_j - \beta_j^0}{S_{b_j}} \tag{4.17}$$

The null is not rejected if the following condition is met:

$$-t_{n-k,\alpha/2} \leq \frac{b_j - \beta_j^0}{S_{b_j}} \leq t_{n-k,\alpha/2} \tag{4.18}$$

Notice the difference between Equation 4.18 and Equation 3.25. The critical value from the t distribution, $t_{n-k,\alpha/2}$, now has $n-k$ degrees of freedom because we are estimating k parameters, rather than just 2 as in the simple regression model. The intuition behind Figures 3.1 and 3.1 still hold. The computer software will also give you the p-value associated with the t test. If the p-value is below your α, you reject the null hypothesis.

For the construction of the confidence intervals we have:

$$1-\alpha = P\big(-t_{n-k,\alpha/2} \leq \tfrac{b_j-\beta_j}{S_{b_j}} \leq t_{n-k,\alpha/2}\big) \tag{4.19}$$

$$1-\alpha = P\big(-t_{n-k,\alpha/2} \cdot S_{b_j} \leq b_j - \beta_j \leq t_{n-k,\alpha/2} \cdot S_{b_j}\big)$$

$$1-\alpha = P\big(b_j - t_{n-k,\alpha/2} \cdot S_{b_j} \leq \beta_j \leq b_j + t_{n-k,\alpha/2} \cdot S_{b_j}\big).$$

4.5 Regression output in Gretl

Gretl is an open-source (free) software package for econometric analysis written in the C programming language. It can be downloaded from:

`http://gretl.sourceforge.net/`

Just follow the instructions to install it in your computer.

Once you loaded the data set in Gretl, to estimate Equation 4.2 you need to go to `Model` → `Ordinary Least Squares`. The regression output is:

```
Model 1: OLS, using observations 1-92
Dependent variable: crimes

             coefficient     std. error    t-ratio   p-value
  ----------------------------------------------------------
  const      2193.34        3918.06        0.5598   0.5770
  pop           0.0652716      0.0106262    6.143    2.30e-08 ***
  unem       -279.291        407.791       -0.6849   0.4952
  officers     15.0406         3.57660      4.205    6.25e-05 ***

Mean dependent var   39663.53   S.D. dependent var   29692.10
Sum squared resid    1.39e+10   S.E. of regression   12548.04
R-squared            0.827293   Adjusted R-squared   0.821405
F(3, 88)             140.5107   P-value(F)           1.90e-33
Log-likelihood      -996.7310   Akaike criterion     2001.462
Schwarz criterion    2011.549   Hannan-Quinn         2005.533

Excluding the constant, p-value was highest for variable 3 (unem)
```

A standard way to present the regression output is:

$$\widehat{\text{crimes}} = \underset{(3918.1)}{2193.34} - \underset{(407.79)}{279.291}\,\text{unem} + \underset{(3.5766)}{15.0406}\,\text{officers} + \underset{(0.010626)}{0.0652716}\,\text{pop}$$

$$N = 92 \quad \bar{R}^2 = 0.8214 \quad F(3,88) = 140.51 \quad \hat{\sigma} = 12548.$$

(standard errors in parentheses)

To obtain the confidence intervals for the coefficients as presented in Equation 4.19 in the Gretl regression output window you need to go to `Analysis` → `Confidence intervals for the coefficients` to obtain:

```
t(88, 0.025) = 1.987

    VARIABLE         COEFFICIENT      95% CONFIDENCE INTERVAL

         const       2193.34           -5592.97       9979.66
           pop          0.0652716      0.0441542    0.0863890
          unem       -279.291          -1089.69       531.107
      officers         15.0406          7.93282       22.1483
```

4.6 Multicollinearity

Multicollinearity is when two explanatory variables are highly correlated. In addition, if their coefficients have a large population variances, we are at risk of getting erratic estimates of the coefficients. There could also be multicollinearity when there is an approximate linear relationship between more than two variables.

A simple test for multicollinearity is based in the Variance Inflation Factors. To implement this text in Gretl, in the regression output window go to `Test` → `Collinearity`:

```
Variance Inflation Factors
```

```
Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

            pop    4.180
           unem    1.094
       officers    4.371

VIF(j) = 1/(1 - R(j)^2), where R(j) is the multiple correlation
coefficient between variable j and the other independent variables

Properties of matrix X'X:

 1-norm = 2.0266173e+013
 Determinant = 6.6859257e+024
 Reciprocal condition number = 4.681685e-013
```

Based on these results, we do not have a multicollinearity problem in the estimation of Equation 4.2.

4.7 Goodness of fit: R^2 and $\bar{R}^2$

The R^2 in multiple regression analysis has the same interpretation as in a simple regression. It is the proportion of the variation in Y explained by the regression model

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \tag{4.20}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{4.21}$$

where $\hat{Y}$ represents the fitted values of the regression equation

$$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3 + \cdots + b_k X_k. \tag{4.22}$$

4.8 F tests

Given the population regression model

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + u, \tag{4.23}$$

we can use the F test to test if all the slope coefficients $\beta_2, \beta_3, \ldots, \beta_k$ are jointly equal to zero. That is, let the null hypothesis be:

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0. \tag{4.24}$$

The alternative hypothesis (H_0) is that at least one of the slope coefficients is different from zero. The multiple regression version of the F statistic is:

$$F_{k-1,n-k} = \frac{ESS/(k-1)}{RSS/(n-k)}. \tag{4.25}$$

The idea is to compare this F statistic to the critical level found in the F distribution tables with $k-1$ and $n-k$ degrees of freedom. Computer software automatically computes this F statistic and the corresponding p-value for the null in Equation 4.22. This F statistic can also be written in terms of the R^2:

$$F_{k-1,n-k} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}. \tag{4.26}$$

Consider the example presented in Section 4.5. The F statistic is 140.5107 with 3 and 88 degrees of freedom and has a corresponding p-value of 0.000. Then, because the p-value is below $\alpha = 5\%$ then we reject the null hypothesis that the slope coefficients on `pop`, `unem`, and `officers` are jointly equal to zero.

4.9 Adjusted R^2, $\bar{R}^2$

One concern with the R^2 is that it will always go up as we include more variables into the model. Hence, it is a poor way to compare models. On the other hand a similar statistic, the adjusted R^2 ($\bar{R}^2$) is built on the R^2 but with the difference that $\bar{R}^2$ penalizes for the loss of the degrees of freedom as we include more variables into the model. Therefore, the $\bar{R}^2$ can either go up or down as we include more variable into the model. It is defined as:

$$\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1-R^2). \tag{4.27}$$

Chapter 5
Transformations of Variables and Interactions

5.1 Basic idea

One limitation in the linear regression analysis is that the dependent variable has to be linear in the parameters:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + u. \tag{5.1}$$

However, there are equations that are not linear, for example:

$$Y = \beta_1 + \beta_2 X_2 + X_3^{\beta_3} + u. \tag{5.2}$$

This Equation 5.2 cannot be estimated using OLS. One way to estimate nonlinear models is by using Nonlinear Least Squares (NLS), which is an extension of the methods we discussed before. In this chapter, rather that focusing on NLS, we will see how transformations in the variables can allow us to use OLS on a variety of nonlinear models. For example, consider the estimation of the following Cobb-Douglas production function:

$$P_i = A L_i^{\beta_2} K_i^{\beta_3} e^{\varepsilon_i}, \tag{5.3}$$

where P_i is total production or total output, A is a technology constant, K_i is the amount of capital, and L_i is labor. Taking natural logs we have:

$$\log P_i = \log A + \beta_2 \log L_i + \beta_3 \log K_i + \varepsilon_i. \tag{5.4}$$

If we simple set $Y_i = \log P_i$, $\beta_1 = \log A$, $X_2 = \log L_i$, and $X_3 = \log K_i$ we can write Equation 5.4 as:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i, \tag{5.5}$$

that can be easily estimated via OLS. β_2 and β_3 will correspond to the ones given in Equation 5.3. Another example of a model that can be estimated with OLS is:

$$Y_i = \beta_1 + \beta_2 Z_{2i}^2 + \beta_3 \sqrt{Z_{3i}} + \beta_4 \frac{1}{Z_{4i}} + \varepsilon_i. \tag{5.6}$$

We just need to replace $X_{2i} = Z_{2i}^2$, $X_{3i} = \sqrt{Z_{3i}}$, $X_{4i} = \frac{1}{Z_{4i}}$.

5.2 Logarithmic transformations

To explain the logarithm transformation let's go over one example in Gretl. If we want to estimate the following model:

$$\log \texttt{crime}_i = \beta_1 + \beta_2 \log \texttt{pop}_i + \beta_3 \texttt{unem}_i + \beta_4 \texttt{offi}_i + u_i, \tag{5.7}$$

you need to create the new variables first. Go to `Add` → `Define new variable` and type:

```
logcrime = log(crime)
```

This will generate the new variable `logcrime`. Do the same thing for log population and then estimate the model. The regression output is:

```
Model 1: OLS, using observations 1-92
Dependent variable: logcrime

             coefficient    std. error    t-ratio    p-value
  ---------------------------------------------------------
  const      -0.709735      0.807193      -0.8793   0.3817
  unem       -0.00456848    0.00903041    -0.5059   0.6142
  offi        0.000144915   6.15429e-05    2.355    0.0208    **
  logpop      0.864044      0.0662782     13.04     2.92e-022 ***

Mean dependent var   10.33774   S.D. dependent var   0.742056
Sum squared resid    6.883563   S.E. of regression   0.279683
R-squared            0.862628   Adjusted R-squared   0.857945
F(3, 88)             184.1989   P-value(F)           8.20e-38
Log-likelihood      -11.28034   Akaike criterion     30.56069
Schwarz criterion    40.64784   Hannan-Quinn         34.63195

Excluding the constant, p-value was highest for variable 3 (unem)
```

$$\widehat{\texttt{logcrime}} = \underset{(0.80719)}{-0.709735} - \underset{(0.00903)}{0.00456}\,\texttt{unem} + \underset{(6.1543\text{e}{-}005)}{0.000144915}\,\texttt{offi} + \underset{(0.0663)}{0.8640}\,\texttt{logpop}$$

$$N = 92 \quad \bar{R}^2 = 0.8579 \quad F(3,88) = 184.20 \quad \hat{\sigma} = 0.27968$$

(standard errors in parentheses)

First, notice how the coefficients are very different from the one obtain with no logarithm transformation. Here the interpretation is different. β_2 is interpreted as the elasticity of `crime` with respect to `pop`:

$$\beta_2 = \frac{\Delta \texttt{crime}/\texttt{crime}}{\Delta \texttt{pop}/\texttt{pop}}. \tag{5.8}$$

A one percentage increase in pop will increase crime by 0.864 percent. Δcrime/crime is interpreted as a percentage change in crime. For β_4 we have:

$$\beta_4 = \frac{\Delta \texttt{crime}/\texttt{crime}}{\Delta \texttt{offi}}. \tag{5.9}$$

Here, a one unit increase in offi is associated with a 0.014% (0.00014×100 percent) increase in crime.

5.3 Quadratic terms

So far we have bee estimating the marginal effects (βs) that are constant across all possible values of X. The simplest way to introduce nonlinearities in the marginal effect is to estimate the model with quadratic terms. For example, let the model be:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i. \tag{5.10}$$

In this case the marginal effect of X on Y is given by:

$$\frac{\Delta Y}{\Delta X} = \beta_2 + 2 \cdot \beta_3 X_i. \tag{5.11}$$

If we want to estimate the marginal effect of experience of wages and in addition we allow for a nonlinear effect we can estimate:

$$\texttt{wage}_i = \beta_1 + \beta_2 \texttt{exper}_i + \beta_3 \texttt{expersq}_i + \varepsilon_i, \tag{5.12}$$

where wage is average hourly earnings, exper is years of experience and expersq is the number of years of experience squared. The Gletl output is the following:

```
Model 1: OLS, using observations 1-526
Dependent variable: wage

             coefficient   std. error    t-ratio    p-value
  ---------------------------------------------------------
  const       3.72541      0.345939      10.77     1.46e-024 ***
  exper       0.298100     0.0409655      7.277    1.26e-012 ***
  expersq    -0.00612989   0.000902517   -6.792    3.02e-011 ***

Mean dependent var   5.896103   S.D. dependent var   3.693086
Sum squared resid    6496.147   S.E. of regression   3.524334
R-squared            0.092769   Adjusted R-squared   0.089300
F(2, 523)            26.73982   P-value(F)           8.77e-12
Log-likelihood      -1407.455   Akaike criterion     2820.910
Schwarz criterion    2833.706   Hannan-Quinn         2825.920
```

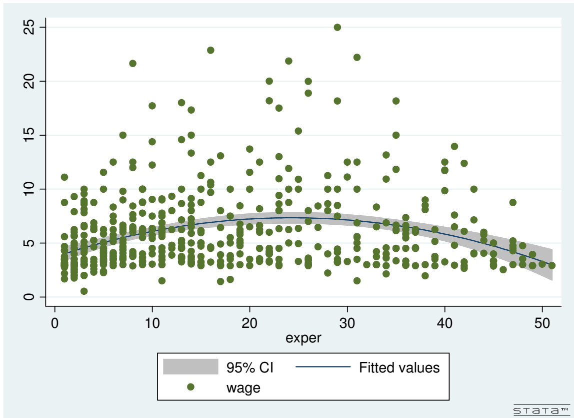

Fig. 5.1 Predicted values for Equation 5.12

$$\widehat{\texttt{wage}} = \underset{(0.34594)}{3.72541} + \underset{(0.040966)}{0.298100}\,\texttt{exper} - \underset{(0.00090252)}{0.00612989}\,\texttt{expersq}$$
$$N = 526 \quad \bar{R}^2 = 0.0893 \quad F(2,523) = 26.740 \quad \hat{\sigma} = 3.5243$$
$$\text{(standard errors in parentheses)}$$

Here, the marginal effect of experience on average hourly wage is:

$$\begin{aligned}\frac{\Delta\texttt{wage}}{\Delta\texttt{exper}} &= 0.2981 + 2\cdot(-0.006)\texttt{exper}\\ &= 0.2981 - 0.012\texttt{exper}.\end{aligned}$$

For a person with 2 years of experience, the effect of an additional year of experience on wage is 0.2741 (=0.2981 - 0.012 $\times$ 2) and for a person with 15 years of experience, the marginal effect of an additional year of experience is 0.1181 (=0.2981 - 0.012 $\times$ 15). Hence, we can say that for a reasonable range of years of experience, experience has a positive effect on wage. In addition, this effect is smaller as you accumulate more experience.

Figure 5.1 show the fitted regression line along with the 95% confidence interval for the fitted values and the actual data. This figure clearly shows the nonlinear marginal effect and innlustrates how wages increase with experience for about the first 25 years, but then wages decrease later on.

5.4 Interaction terms

A second popular approach to allow for the marginal effect to change over different values of X is to include interaction terms in the regression equation. For example,

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 (X_2 \times X_3) + \varepsilon_i. \tag{5.13}$$

In this case the marginal effect of X_2 on Y depends on X_3 is given by:

$$\frac{\Delta Y}{\Delta X_2} = \beta_2 + \beta_3 X_3. \tag{5.14}$$

Consider the next example with the interaction between `exper` and `educ` in a wage equation:

$$\texttt{wage}_i = \beta_1 + \beta_2 \texttt{exper}_i + \beta_3 (\texttt{exper}_i \times \texttt{educ}_i) + \varepsilon_i, \tag{5.15}$$

where the marginal effect of experience on wage depends on the level of education:

$$\frac{\Delta \texttt{wage}}{\Delta \texttt{exper}} = \beta_2 + \beta_3 \texttt{educ}. \tag{5.16}$$

When estimating this equation in Gretl we have to make sure we generate the interaction term first. That is, go to `Add` → `Define new variable` and type:

```
expereduc = exper*educ
```

Then we are ready to estimate the equation via OLS. The regression output is:

```
Model 1: OLS, using observations 1-526
Dependent variable: wage

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const        4.88993      0.242730     20.15     3.11e-067 ***
  exper       -0.188124     0.0253904    -7.409    5.13e-013 ***
  expereduc    0.0207731    0.00217625    9.545    5.17e-020 ***

Mean dependent var   5.896103   S.D. dependent var   3.693086
Sum squared resid    6020.313   S.E. of regression   3.392803
R-squared            0.159223   Adjusted R-squared   0.156008
F(2, 523)            49.52175   P-value(F)           2.01e-20
Log-likelihood      -1387.449   Akaike criterion     2780.897
Schwarz criterion    2793.693   Hannan-Quinn         2785.908
```

$$\widehat{\texttt{wage}} = \underset{(0.24273)}{4.88993} - \underset{(0.025390)}{0.188124}\,\texttt{exper} + \underset{(0.0021762)}{0.0207731}\,\texttt{expereduc}$$

$$N = 526 \quad \bar{R}^2 = 0.1560 \quad F(2,523) = 49.522 \quad \hat{\sigma} = 3.3928$$

(standard errors in parentheses)

Here, the marginal effect of experience on wage is:

$$\frac{\Delta \texttt{wage}}{\Delta \texttt{exper}} = -0.1881 + 0.0208\texttt{educ}$$

For a person with twelve years education (high school), the marginal effect from an additional year of education is 0.0615 (=-0.1881+0.208$\times$12). However, with more education the marginal effect is larger. A person with 16 years of education (high school + college) will have a marginal effect of 0.1447 (=-0.1881+0.208$\times$16). Notice that for an important range of education the marginal effect is positive, meaning that more experience leads to higher wages. In addition, the effect if larger if you have more education. This means that going to school is not only good because it directly increases your expected wage but also makes additional years of experience more valuable.

Chapter 6
Analysis with Qualitative Information: Dummy Variables

In previous chapters, the dependent and the independent variables in our regression equations had a *quantitative* meaning. That is, the magnitude of the variable had a useful information, for example, years of education, years of experience, unemployment rate, or wage. In this chapter we will analyze how to introduce *qualitative* information into a regression equation. Example of qualitative information includes marital status, gender, race, industry (manufacturing, retail, etc.) or geographical region (south, north, west, etc.).

6.1 Describing qualitative information

Qualitative factors often come in the form of binary information: a person is female of male; a person does or does not own a computer; a person is married or not. In all these cases the relevant information can be captured by a binary variable, also called a dummy variable or zero-one variable. In defining a dummy variable we must decide which event is assigned a value of one and which a value of zero. Table 6.1 shows how two dummy variables (`female` and `married`) look in the data set.

Table 6.1 A partial Listing of the Data in Wage.xls

`person`	`wage`	`educ`	`exper`	`female`	`married`
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	5	1	0

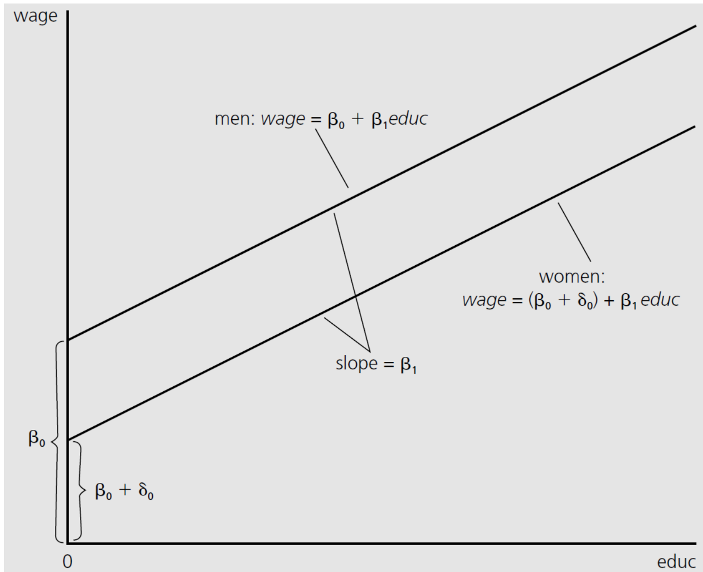

Fig. 6.1 Graph of $\texttt{wage} = \beta_0 + \delta_0 \texttt{female} + \beta_1 \texttt{educ}$ for $\delta_0 < 0$.

6.2 A single dummy independent variable

The simplest case is when we have a single dummy independent variable. Let's consider the following model:

$$\texttt{wage} = \beta_0 + \delta_0 \texttt{female} + \beta_1 \texttt{educ} + \varepsilon \tag{6.1}$$

We use the parameter δ_0 to emphasize the fact that `female` corresponds to a dummy variable. If the person is a female we have `female` = 1, and if the person is a male, we have `female` = 0. The parameter δ_0 has the following interpretation: δ_0 is the difference in hourly wage between females and males, given the same amount of education (and the error term ε). Thus, the coefficient δ_0 determines whether there is discrimination against women: if $\delta_0 < 0$, it means that on average, women earn less than men.

The interpretation of δ_0 (when $\delta < 0$) can be depicted graphically in Figure 6.1 as an *intercept shift* between males an females.

Let's estimate the following more interesting model:

$$\texttt{wage} = \beta_0 + \delta_0 \texttt{female} + \beta_1 \texttt{educ} + \beta_2 \texttt{exper} + \beta_1 \texttt{tenure} + \varepsilon \tag{6.2}$$

The regression output in Gretl is:

```
Model 2: OLS, using observations 1-526
Dependent variable: wage
```

```
             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const      -1.56794      0.724551     -2.164    0.0309    **
  female     -1.81085      0.264825     -6.838    2.26e-011 ***
  educ        0.571505     0.0493373    11.58     9.09e-028 ***
  exper       0.0253959    0.0115694     2.195    0.0286    **
  tenure      0.141005     0.0211617     6.663    6.83e-011 ***

Mean dependent var   5.896103   S.D. dependent var   3.693086
Sum squared resid    4557.308   S.E. of regression   2.957572
R-squared            0.363541   Adjusted R-squared   0.358655
F(4, 521)            74.39801   P-value(F)           7.30e-50
Log-likelihood      -1314.228   Akaike criterion     2638.455
Schwarz criterion    2659.782   Hannan-Quinn         2646.805
```

$$\widehat{\texttt{wage}} = \underset{(0.72455)}{-1.56794} - \underset{(0.26483)}{1.81085}\,\texttt{female} + \underset{(0.049337)}{0.571505}\,\texttt{educ} + \underset{(0.011569)}{0.0253959}\,\texttt{exper}$$
$$+ \underset{(0.021162)}{0.141005}\,\texttt{tenure}$$
$$N = 526 \quad \bar{R}^2 = 0.3587 \quad F(4,521) = 74.398 \quad \hat{\sigma} = 2.9576$$
(standard errors in parentheses)

Where it is easy to see that $\delta_0 = -1.81$. If we want to test the null hypothesis that there is no difference between men and women, $H_0 : \delta_0 = 0$. The alternative hypothesis is that there is discrimination against women, $H_1 : \delta_0 < 0$. Based on the p-value we reject the null and conclude that there is discrimination, females make two dollars and twenty seven cents less per hour than males. This is after controlling for differences in education, experience and tenure.

It is illustrative to additionally estimate the following equation:

$$\texttt{wage} = \beta_0 + \delta_0\texttt{female} + \varepsilon \tag{6.3}$$

where we do not control for education, experience or tenure. The regression output is:

```
Model 3: OLS, using observations 1-526
Dependent variable: wage

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const        7.09949      0.210008    33.81     8.97e-134 ***
  female      -2.51183      0.303409    -8.279    1.04e-015 ***

Mean dependent var   5.896103   S.D. dependent var   3.693086
Sum squared resid    6332.194   S.E. of regression   3.476254
R-squared            0.115667   Adjusted R-squared   0.113979
F(1, 524)            68.53668   P-value(F)           1.04e-15
Log-likelihood      -1400.732   Akaike criterion     2805.464
Schwarz criterion    2813.995   Hannan-Quinn         2808.804
```

$$\widehat{\texttt{wage}} = \underset{(0.21001)}{7.09949} - \underset{(0.30341)}{2.51183}\,\texttt{female}$$
$$N = 526 \quad \bar{R}^2 = 0.1140 \quad F(1,524) = 68.537 \quad \hat{\sigma} = 3.4763$$
$$\text{(standard errors in parentheses)}$$

The expected (predicted) wage for females is $\widehat{\texttt{wage}} = 7.099 - 2.512\texttt{1} = 4.587$, while the expected wage for males is $\widehat{\texttt{wage}} = 7.099 - 2.512\texttt{0} = 7.099$. This is not controlling for differences in education, experience or tenure. Once we control for those differences, the wage gap between these two groups is smaller and equal to $\delta_0 = -1.81$.

What is the interpretation of the coefficient on a dummy variable if the dependent variable is in logs? Here the coefficient has a *percentage* interpretation. Let's say we want to estimate the following equation:

$$\log \texttt{wage} = \beta_0 + \delta_0 \texttt{female} + \beta_1 \texttt{educ} + \beta_2 \texttt{exper} + \beta_3 \texttt{tenure} + \varepsilon \tag{6.4}$$

that has the following Gretl estimation output:

$$\widehat{\texttt{logwage}} = \underset{(0.10190)}{0.501348} - \underset{(0.037246)}{0.301146}\,\texttt{female} + \underset{(0.0069389)}{0.0874623}\,\texttt{educ} + \underset{(0.0016271)}{0.00462938}\,\texttt{exper}$$
$$+ \underset{(0.0029762)}{0.0173670}\,\texttt{tenure}$$
$$N = 526 \quad \bar{R}^2 = 0.3876 \quad F(4,521) = 84.072 \quad \hat{\sigma} = 0.41596$$
$$\text{(standard errors in parentheses)}$$

The coefficient on `female`, δ_0, implies that for the same levels of education, experience, and tenure, women earn approximately 100(0.301) = 30.1% less than men.

6.3 Dummy variables for multiple categories

One can use several dummy variables in the same equation. For example, we can add the dummy variable `married` to Equation 6.3 to obtain:

$$\texttt{wage} = \beta_0 + \delta_0 \texttt{female} + \delta_1 \texttt{married} + \varepsilon \tag{6.5}$$

In Gretl we have,

$$\widehat{\texttt{wage}} = \underset{(0.29634)}{6.18043} - \underset{(0.30261)}{2.29440}\,\texttt{female} + \underset{(0.30971)}{1.33948}\,\texttt{married}$$
$$N = 526 \quad \bar{R}^2 = 0.1429 \quad F(2,523) = 44.779 \quad \hat{\sigma} = 3.4190$$
$$\text{(standard errors in parentheses)}$$

The coefficient on `married` gives the (approximate) difference in wages between married and non married individuals. Based on these results, married individuals have higher hourly wages. On important restriction in Equation 6.5 is that it restricts the effect of marital status on wages is the same whether you are male of female. If we are interested in this difference we can estimate an alternative model with additional categories. In particular we need four categories: (1) married men, (2) married women (3) single men, and (4) single woman. We must select a base group (for example, single men) and create the dummy variables for the other three groups.

$$\begin{aligned}\texttt{marrmale} &= \texttt{married} \times (1 - \texttt{female}) \\ \texttt{marrfem} &= \texttt{married} \times \texttt{female} \\ \texttt{singfem} &= (1 - \texttt{married}) \times \texttt{female}\end{aligned}$$

The equation we want to estimate is:

$$\log \texttt{wage} = \beta_0 + \delta_0 \texttt{marrmale} + \delta_1 \texttt{marrfem} + \delta_2 \texttt{singfem} + \varepsilon \tag{6.6}$$

and the estimation output is:

$$\widehat{\texttt{logwage}} = \underset{(0.050987)}{1.5201} + \underset{(0.061554)}{0.4267}\ \texttt{marrmale} - \underset{(0.065524)}{0.0797}\ \texttt{marrfem} - \underset{(0.066804)}{0.1316}\ \texttt{singfem}$$
$$N = 526 \quad \bar{R}^2 = 0.2087 \quad F(3,522) = 47.149 \quad \hat{\sigma} = 0.47284$$

(standard errors in parentheses)

The interpretation of each of the δ coefficients is with respect to the base group. For example $\delta_2 = 0.1316$ means that single females earn approximately 13.16% lower hourly wages than single men (the base group).

6.4 Incorporating ordinal information

Suppose we want to estimate the effect of city credit ratings on the municipal bond interest rate (`MBR`). The credit rating (`CR`) is an ordinal variable and suppose it goes from zero (worst credit) to four (best credit). Under these consideration, a potential candidate for our model is:

$$\texttt{MBR} = \beta_0 + \beta_1 \texttt{CR} + other factors + \varepsilon \tag{6.7}$$

where $other factors$ are just other variables in the model. On concern with this specification is that it is hard to interpret one unit increase in CR. It is easy to talk about an additional year of education or an additional year of income, but credit ratings usually have only an ordinal meaning. Moreover, it is restrictive to assume that each additional unit increase in `CR` has the same effect on `MBR`. An alternative approach is to create separate dummy variables for each of the values of `CR`, that is,

$$
\begin{aligned}
\mathtt{CR}_1 &= 1 \text{ if } \mathtt{CR} = 1 \\
&= 0 \text{ otherwise.} \\
\mathtt{CR}_2 &= 1 \text{ if } \mathtt{CR} = 2 \\
&= 0 \text{ otherwise.} \\
\mathtt{CR}_3 &= 1 \text{ if } \mathtt{CR} = 3 \\
&= 0 \text{ otherwise.} \\
\mathtt{CR}_4 &= 1 \text{ if } \mathtt{CR} = 4 \\
&= 0 \text{ otherwise.}
\end{aligned}
$$

Then we can focus on estimating the following model:

$$\mathtt{MBR} = \beta_0 + \delta_1 \mathtt{CR}_1 + \delta_2 \mathtt{CR}_2 + \delta_3 \mathtt{CR}_3 + other factors + \varepsilon \tag{6.8}$$

Again, we omit one category ($\mathtt{CR}_4$) and the interpretation of the dummy coefficients is relative to the omitted category. For example, δ_2 represents the difference in municipal bond interest rate between ratings $\mathtt{CR}_2$ and $\mathtt{CR}_4$.

6.5 Interactions involving dummy variables

Just as quantitative variables can have interactions, so can dummy variables. Actually, we revisit the estimation of Equation 6.6 to see that the same results can be achieved by including the interaction term between `female` and `married`. The model we want to estimate is:

$$\log \mathtt{wage} = \beta_0 + \delta_0 \mathtt{female} + \delta_1 \mathtt{married} + \delta_2 (\mathtt{female} \times \mathtt{married}) + \varepsilon \tag{6.9}$$

Estimating in Gretl we have:

$$
\begin{aligned}
\widehat{\mathtt{logwage}} = \underset{(0.050987)}{1.5201} - \underset{(0.066804)}{0.1316}\, \mathtt{female} + \underset{(0.061554)}{0.4267}\, \mathtt{married} \\
- \underset{(0.085708)}{0.3748}\, \mathtt{female} \times \mathtt{married}
\end{aligned}
$$

$$N = 526 \quad \bar{R}^2 = 0.2087 \quad F(3, 522) = 47.149 \quad \hat{\sigma} = 0.47284$$

(standard errors in parentheses)

Notice that this regression output is equivalent as the one obtained from Equation 6.6.

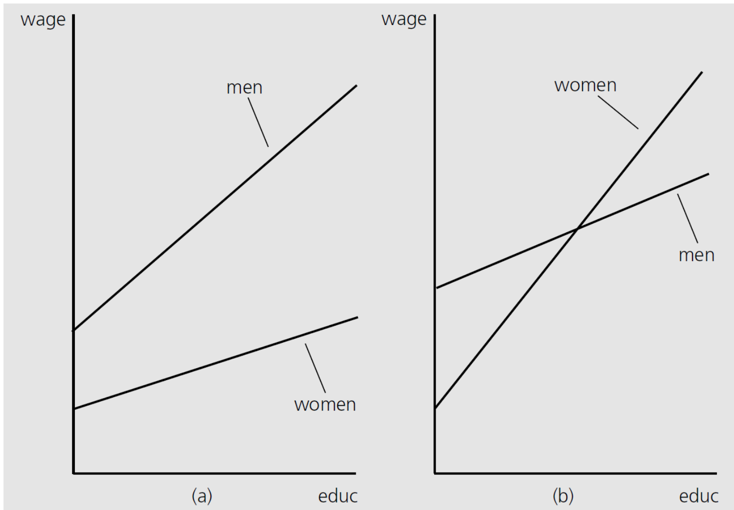

Fig. 6.2 Graph of $\texttt{wage} = \beta_0 + \delta_0 \texttt{female} + \beta_1 \texttt{educ} + \delta_1 \texttt{educ} \times \texttt{female}$.

6.5.1 Allowing for different slopes

Consider the case where we want to estimate the effect of education on hourly wage and in addition, we want for the marginal effect to change based on your gender. This can be done by interacting the `educ` with `female` and estimating the following model:

$$\texttt{wage} = \beta_0 + \delta_0 \texttt{female} + \beta_1 \texttt{educ} + \delta_1 (\texttt{female} \times \texttt{educ}) + \varepsilon \tag{6.10}$$

A graphical approach to this problem in presented in Figure 6.2. The output in Gretl is

$$\widehat{\texttt{wage}} = \underset{(0.84356)}{0.200496} - \underset{(1.3250)}{1.19852}\,\texttt{female} + \underset{(0.064223)}{0.539476}\,\texttt{educ}$$
$$- \underset{(0.10364)}{0.0859990}\,\texttt{female} \times \texttt{educ}$$
$$T = 526 \quad \bar{R}^2 = 0.2555 \quad F(3,522) = 61.070 \quad \hat{\sigma} = 3.1865$$
$$\text{(standard errors in parentheses)}$$

6.5.2 Testing for differences in regression functions across groups

So far we saw that interacting a dummy variable with other independent variables is a powerful tool. Now, we can use this tool to test the null hypothesis that two groups follow the same regression function, against the alternative that one or more of the slopes differs across the two groups. Suppose we want to test whether the same regression model describe college GPA for males and for females. The model is

$$\texttt{cumgpa} = \beta_0 + \beta_1 \texttt{sat} + \beta_2 \texttt{hsperc} + \beta_3 \texttt{tothrs} + \varepsilon, \tag{6.11}$$

where `cumgpa` is cumulative college GPA, `sat` is the SAT score, `hsperc` is the high school rank percentile, and `tothrs` is the total hours of college courses. The regression results in Gretl are

$$\widehat{\texttt{cumgpa}} = \underset{(0.22855)}{0.929111} + \underset{(0.000208)}{0.0009028}\,\texttt{sat} - \underset{(0.00157)}{0.006379}\,\texttt{hsperc} + \underset{(0.000931)}{0.01198}\,\texttt{tothrs}$$

$$N = 732 \quad \bar{R}^2 = 0.2323 \quad F(3,728) = 74.717 \quad \hat{\sigma} = 0.86711$$

$$\text{(standard errors in parentheses)}$$

To allow for a difference in the intercept we just need to include the dummy variable `female`. Then, to allow differences in the slope parameters we need to include interaction terms for each of the variables and `female`. That is

$$\begin{aligned} \texttt{cumgpa} = {} & \beta_0 + \delta_0 \texttt{female} + \beta_1 \texttt{sat} + \delta_1 \texttt{sat} \cdot \texttt{female} \\ & + \beta_2 \texttt{hsperc} + \delta_2 \texttt{hsperc} \cdot \texttt{female} \\ & + \beta_3 \texttt{tothrs} + \delta_3 \texttt{tothrs} \cdot \texttt{female} + \varepsilon \end{aligned} \tag{6.12}$$

The parameter δ_0 is the difference in the intercepts between females and males, δ_1 is the slope difference with respect to `sat` between females and males, and so on. The null hypothesis that `cumgpa` follows the same model for females and males is

$$H_0 : \delta_0 = \delta_1 = \delta_2 = \delta_3 = 0 \tag{6.13}$$

If at least one of the δ_j is different from zero, then the model is different for men and women. After creating the interaction terms, the estimated model in Gretl is

```
Model 2: OLS, using observations 1-732
Dependent variable: cumgpa

                 coefficient    std. error    t-ratio     p-value
  ---------------------------------------------------------------
  const           1.21398       0.264828       4.584      5.37e-06  ***
  sat             0.000611312   0.000235026    2.601      0.0095    ***
  hsperc         -0.00596745    0.00177646    -3.359      0.0008    ***
  tothrs          0.0103004     0.00109284     9.425      5.65e-020 ***
  female         -1.11364       0.528539      -2.107      0.0355    **
  satfemale       0.00111674    0.000500034    2.233      0.0258    **
  hspercfemale    5.07597e-05   0.00410253     0.01237    0.9901
  tothrsfemale    0.00555989    0.00206958     2.686      0.0074    ***
```

```
Mean dependent var   2.080861   S.D. dependent var   0.989617
Sum squared resid    534.3092   S.E. of regression   0.859067
R-squared            0.253652   Adjusted R-squared   0.246436
F(7, 724)            35.15106   P-value(F)           2.54e-42
Log-likelihood      -923.4440   Akaike criterion     1862.888
Schwarz criterion    1899.654   Hannan-Quinn         1877.071
```

$$\widehat{\texttt{cumgpa}} = \underset{(0.26483)}{1.21398} + \underset{(0.00023503)}{0.000611312}\,\texttt{sat} - \underset{(0.0017765)}{0.00596745}\,\texttt{hsperc} + \underset{(0.0010928)}{0.0103004}\,\texttt{tothrs}$$
$$- \underset{(0.52854)}{1.11364}\,\texttt{female} + \underset{(0.00050003)}{0.00111674}\,\texttt{satfemale} + \underset{(0.0041025)}{5.07597\text{e–}005}\,\texttt{hspercfemale}$$
$$+ \underset{(0.0020696)}{0.00555989}\,\texttt{tothrsfemale}$$

$$N = 732 \quad \bar{R}^2 = 0.2464 \quad F(7,724) = 35.151 \quad \hat{\sigma} = 0.85907$$
(standard errors in parentheses)

Now, to test the null hypothesis presented in Equation 6.13 from the window that shows the regression output, we need to go to `Tests` → `Omit variables` and a new window will open. We then have to select the variables to omit. There are `female`, `satfemale`, `hspercfemale`, and `tothrsfemale`. This will estimate the restricted model and the comparison between the restricted model (Equation 6.11) and the full model (Equation 6.12),

```
Model 3: OLS, using observations 1-732
Dependent variable: cumgpa

             coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const       0.929111      0.228552       4.065    5.32e-05  ***
  sat         0.000902834   0.000207870    4.343    1.60e-05  ***
  hsperc     -0.00637913    0.00156785    -4.069    5.24e-05  ***
  tothrs      0.0119779     0.000931383   12.86     2.96e-034 ***

Mean dependent var   2.080861   S.D. dependent var   0.989617
Sum squared resid    547.3649   S.E. of regression   0.867107
R-squared            0.235416   Adjusted R-squared   0.232265
F(3, 728)            74.71707   P-value(F)           3.87e-42
Log-likelihood      -932.2797   Akaike criterion     1872.559
Schwarz criterion    1890.942   Hannan-Quinn         1879.651

Comparison of Model 2 and Model 3:

  Null hypothesis: the regression parameters are zero for the variables
    female, satfemale, hspercfemale, tothrsfemale

  Test statistic: F(4, 724) = 4.4227, with p-value = 0.00154347
  Of the 3 model selection statistics, 1 has improved.
```

The F statistics that Gretl is reporting comes from

$$F = \frac{RSS - RSS_{\mathrm{UR}}}{RSS_{\mathrm{UR}}} \cdot \frac{n-2k}{q}, \tag{6.14}$$

where RSS is the residual sum of squares of the model estimates in Equation 6.11 and $\mathrm{RSS_{UR}}$ is the unrestricted model in Equation 6.12. n is the sample size, k is the number of parameters we are estimating, and q is the number of restrictions when comparing the model in Equation 6.11 and in Equation 6.12. Substituting the values we obtain,

$$F = \frac{547.3649 - 534.3092}{534.3092} \cdot \frac{732 - 2 \cdot 4}{4} = 0.024434 \cdot 181 = 4.4227, \tag{6.15}$$

An alternative way to calculate this F statistic is to follow the formula,

$$F = \frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \cdot \frac{n-2k}{k}, \tag{6.16}$$

where RSS is the residual sum of squares of the model estimates in Equation 6.11. $\mathrm{RSS_1}$ and $\mathrm{RSS_2}$ are the residual sum of squares of the model estimated in Equation 6.11 using only the females in the sample ($\mathrm{RSS_1}$) and using only the males in the sample ($\mathrm{RSS_2}$). As before, n is the sample size and k is the number of parameters we are estimating. The estimation of Equation 6.11 with just females is:

```
Model 5: OLS, using observations 1-180
Dependent variable: cumgpa

             coefficient   std. error    t-ratio    p-value
  ----------------------------------------------------------
  const       0.100346     0.481095       0.2086   0.8350
  sat         0.00172805   0.000464216    3.723    0.0003    ***
  hsperc     -0.00591669   0.00388949    -1.521    0.1300
  tothrs      0.0158603    0.00184854     8.580    4.82e-015 ***

Mean dependent var   2.268611   S.D. dependent var   1.126549
Sum squared resid    143.6897   S.E. of regression   0.903559
R-squared            0.367483   Adjusted R-squared   0.356702
F(3, 176)            34.08447   P-value(F)           2.03e-17
Log-likelihood      -235.1319   Akaike criterion     478.2638
Schwarz criterion    491.0356   Hannan-Quinn         483.4422
```

and with just males is:

```
Model 6: OLS, using observations 1-552
Dependent variable: cumgpa

             coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const       1.21398       0.260270       4.664    3.90e-06  ***
  sat         0.000611312   0.000230981    2.647    0.0084    ***
  hsperc     -0.00596745    0.00174588    -3.418    0.0007    ***
  tothrs      0.0103004     0.00107403     9.590    3.06e-020 ***

Mean dependent var   2.019638   S.D. dependent var   0.933655
```

```
Sum squared resid    390.6194   S.E. of regression   0.844280
R-squared            0.186740   Adjusted R-squared   0.182288
F(3, 548)            41.94377   P-value(F)           2.06e-24
Log-likelihood      -687.8093   Akaike criterion     1383.619
Schwarz criterion    1400.873   Hannan-Quinn         1390.360
```

Using the formula in Equation 6.17,

$$F = \frac{547.3649-(143.6897+390.6194)}{143.6897+390.6194}\cdot\frac{732-2\cdot 4}{4} = 0.024434\cdot 181 = 4.4227, \tag{6.17}$$

which is the same result as in Equation 6.15. This version of the F test is know also as the *Cho* test. A large F statistic is evidence against the null hypothesis. In our example the F statistic of 4.4227 has an associated p-value of 0.0015, below the usual 0.05 (or 5%). Hence, we reject the null hypothesis that there is no difference between the equation for females and the equation for males. This means that there is difference and we are better off estimating Equation 6.12 instead of Equation 6.11.

The key to estimate Equation 6.11 with just the female portion of the data change the sample. To do this go to `Sample` → `Restrict, based on criterion...`, then after a new window shows up, select the "use dummy variable" and then `female`. Once the sample is restricted, just estimate the model using Ordinary Least Squares again.

6.6 The dummy variable trap

The dummy variable trap occurs when there is an exact linear relationship among the variables in the regression model. That is the reason why we do not include `female` and `male` in the same regression equation because `female` + `male` = 1. The same occurs when we have more than one category and we should always omit one of the categories (base group). Than is why `singmen` does not appear in Equation 6.6 (`marrmale` + `singmale` + `marrfem` + `singfem` = 1).

Chapter 7
Specification of Regression Variables

So far we assumed we know what are the variables that needed to be in our regression model. However, what happens if we include in the regression model a variable that should not be there? What if we leave out a variable that should be included? Can we a proxy for a variable that we do not observe? These are the main question this chapter will address.

7.1 Model specification

What happens in practice is that it is difficult to be sure about the correct specification of the regression model. While theory may help, it usually depends on simplifying assumptions that may not necessarily hold. The properties of the regression estimates depend crucially on the validity of the specification of the model. The following is a quick summary of the consequences of misspecifying the regression model:

1. If you leave out a variable that should be included. The regression estimates are potentially biased. The standard errors of the coefficients and the corresponding t and F tests are in general invalid.
2. If you include a variable that should not be in the model. The coefficients will not be biased, but they are potentially inefficient.

7.2 Omitting a variable

7.2.1 The bias problem

Suppose that the true regression model that we should be estimated is given by

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u. \tag{7.1}$$

However, we do not have the variable X_3 or maybe we have it but we do not include it in the model. Hence, we estimate the following model

$$Y = \beta_1 + \beta_2 X_2 + u. \tag{7.2}$$

Then the predicted or fitted values are

$$\hat{Y} = b_1 + b_2 X_2 \tag{7.3}$$

Recall from previous chapters that the formula to estimate b_2 is given by

$$b_2 = \frac{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2} \tag{7.4}$$

We say that b_2 is unbiased if its expected value is equal to the true population parameter β_2. If we plug Equation 7.1 into Equation 7.4 and take expectations we obtain

$$\begin{aligned} E[b_2] &= E\Big[\frac{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2}\Big] \\ &= \beta_2 + \beta_3 \frac{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)}{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2}. \end{aligned} \tag{7.5}$$

For b_2 to be unbiased we need that the second term on the right-hand side be equal to zero. This term is known as the *omitted variable bias* and it will be zero if $\beta_3 = 0$ or if $\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)/\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2$ is equal to zero. Then the conditions for b_2 to be unbiased in the estimation of Equation 7.2 are:

1. That X_3 does not affect Y. That is, $\beta_3 = 0$.
2. That X_2 and X_3 are linearly uncorrelated. That is, the slope coefficient when we regress X_3 on X_2 os zero, $\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)]/[\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2 = 0$.

7.2.2 Invalid statistical tests

When a variable is omitted from the model, the standard errors of the coefficients and the texts statistics are generally invalid. This means that the t and F tests cannot be used.

7.2.3 Example

Consider the case where the true model to explain wages is given by

$$\texttt{wage} = \beta_1 + \beta_2\texttt{educ} + \beta_3\texttt{ability} + u. \tag{7.6}$$

That is, your `wage` is determined by your number of years of formal education (`educ`) and your `ability`. The problem in this equation is that actually it is very difficult to measure `ability`. Hence, we decide omit it and estimate the following model

$$\texttt{wage} = \beta_1 + \beta_2\texttt{educ} + u. \tag{7.7}$$

What is the problem with the estimate of β_2 if we use Equation 7.7? Is is biased! To get an idea of the size of the bias we will proxy `ability` with another variable, `IQ`. Equation 7.5 becomes

$$E[b_2] = \beta_2 + \beta_3 \frac{\sum_{i=1}^{n}(\texttt{educ}_i - \overline{\texttt{educ}})(\texttt{IQ}_i - \overline{\texttt{IQ}})}{\sum_{i=1}^{n}(\texttt{educ}_i - \overline{\texttt{educ}})^2}. \tag{7.8}$$

Notice that we can actually analyze if the bias is positive or negative based on the signs of the second part on the right-hand size. It seems that β_3 should be positive because higher ability (or IQ) should be correlated positively with wages. Moreover, the part that multiplies β_3 should also be positive because education and ability (or IQ) seem to be positively correlated. Hence, the whole second part on the right-hand side is positive, implying that β_2 is biased upwards. This means that on average we will be getting a larger coefficient (by estimating Equation 7.7) than the true coefficient (if we were estimating the true Equation 7.6).

Let's look at this empirically by estimating Equations 7.6 and 7.7 with real data (where we use `IQ` in place of `ability`):

```
Model 1: OLS, using observations 1-935
Dependent variable: wage

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       146.952       77.7150      1.891     0.0589    *
  educ         60.2143       5.69498    10.57      9.35e-025 ***

Mean dependent var   957.9455   S.D. dependent var   404.3608
Sum squared resid    1.36e+08   S.E. of regression   382.3203
R-squared            0.107000   Adjusted R-squared   0.106043
F(1, 933)            111.7929   P-value(F)           9.35e-25
Log-likelihood      -6885.458   Akaike criterion     13774.92
Schwarz criterion    13784.60   Hannan-Quinn         13778.61
```

$$\widehat{\texttt{wage}} = \underset{(77.715)}{146.952} + \underset{(5.6950)}{60.2143}\,\texttt{educ}$$

$$N = 935 \quad \bar{R}^2 = 0.1060 \quad F(1,933) = 111.79 \quad \hat{\sigma} = 382.32$$

(standard errors in parentheses)

```
Model 2: OLS, using observations 1-935
Dependent variable: wage

             coefficient   std. error   t-ratio    p-value
```

```
  ---------------------------------------------------------
  const      -128.890      92.1823      -1.398    0.1624
  educ         42.0576      6.54984      6.421    2.15e-010 ***
  IQ            5.13796    0.955827     5.375    9.66e-08  ***

Mean dependent var   957.9455   S.D. dependent var   404.3608
Sum squared resid    1.32e+08   S.E. of regression   376.7300
R-squared            0.133853   Adjusted R-squared   0.131995
F(2, 932)            72.01515   P-value(F)           8.27e-30
Log-likelihood      -6871.185   Akaike criterion     13748.37
Schwarz criterion    13762.89   Hannan-Quinn         13753.91
```

$$\widehat{\texttt{wage}} = \underset{(92.182)}{-128.890} + \underset{(6.5498)}{42.0576}\,\texttt{educ} + \underset{(0.95583)}{5.13796}\,\texttt{IQ}$$

$$N = 935 \quad \bar{R}^2 = 0.1320 \quad F(2,932) = 72.015 \quad \hat{\sigma} = 376.73$$

(standard errors in parentheses)

The empirical results are consistent with our theoretical analysis. The estimate of β_2 in Equation 7.7 is too large (upward biased). The bias can be obtained separately by estimating a regression of `IQ` on `educ` and then plugging the results in Equation 7.8

```
Model 3: OLS, using observations 1-935
Dependent variable: IQ

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       53.6872       2.62293      20.47    3.36e-077 ***
  educ         3.53383      0.192210     18.39    1.16e-064 ***

Mean dependent var   101.2824   S.D. dependent var   15.05264
Sum squared resid    155346.5   S.E. of regression   12.90357
R-squared            0.265943   Adjusted R-squared   0.265157
F(1, 933)            338.0192   P-value(F)           1.16e-64
Log-likelihood      -3716.973   Akaike criterion     7437.946
Schwarz criterion    7447.627   Hannan-Quinn         7441.637
```

$$\widehat{\texttt{IQ}} = \underset{(2.6229)}{53.6872} + \underset{(0.19221)}{3.53383}\,\texttt{educ}$$

$$N = 935 \quad \bar{R}^2 = 0.2652 \quad F(1,933) = 338.02 \quad \hat{\sigma} = 12.904$$

(standard errors in parentheses)

Replacing the valued in Equation 7.8

$$\begin{aligned} E[b_2] &= \beta_2 + \beta_3 \frac{\sum_{i=1}^{n}(\texttt{educ}_i - \overline{\texttt{educ}})(\texttt{IQ}_i - \overline{\texttt{IQ}})}{\sum_{i=1}^{n}(\texttt{educ}_i - \overline{\texttt{educ}})^2}. \\ &= \beta_2 + 5.13796 \times 3.53383 \\ &= \beta_2 + 18.15667 \end{aligned} \tag{7.9}$$

That is exactly the difference between the coefficients in Equations 7.6 and 7.7, 60.2143 - 42.0576 = 18.15667.

7.3 Including a variable that should not be included

Suppose that the true population model is given by

$$Y = \beta_1 + \beta_2 X_2 + u. \tag{7.10}$$

However, for some season you include X_3 and end up estimating the following model

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u. \tag{7.11}$$

In a regression model like Equation 7.11 with two variables (X_2 and X_3) the OLS estimator for b_2 is given by

$$\begin{aligned} b_2 = & \frac{\sum_{i=1}^n (X_{2i} - \bar{X}_2)(Y_i - \bar{Y}) \sum_{i=1}^n (X_{3i} - \bar{X}_3)^2}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2 \sum_{i=1}^n (X_{3i} - \bar{X}_3)^2 - \left(\sum_{i=1}^n (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)\right)^2} \\ & - \frac{\sum_{i=1}^n (X_{3i} - \bar{X}_3)(Y_i - \bar{Y}) \sum_{i=1}^n (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2 \sum_{i=1}^n (X_{3i} - \bar{X}_3)^2 - \left(\sum_{i=1}^n (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)\right)^2} \end{aligned} \tag{7.12}$$

Which is certainly different from the OLS estimator for b_2 in Equation 7.10,

$$b_2 = \frac{\sum_{i=1}^n (X_{2i} - \bar{X}_2)(Y_i - \bar{Y})}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2} \tag{7.13}$$

Interestingly, b_2 in both Equations (7.12 and 7.13) is unbiased, $E(b_2) = \beta_2$. Hence, estimating the effect of X_2 on Y will yield unbiased estimates even if we include irrelevant variables. Then, what is the problem? Including irrelevant variables will inflate the standard errors of the coefficients. This means that the estimate b_2 from Equation 7.11 will be inefficient. The implied population variance of b_2 in Equation 7.11 is

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2} \cdot \frac{1}{(1 - r_{X_2X_3}^2)} \tag{7.14}$$

where $r_{X_2X_3}^2$ is the correlation coefficient between X_2 and X_3, while the population variance of b_2 in Equation 7.10 is

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2}. \tag{7.15}$$

Notice that because $0 \leq r_{X_2X_3}^2 \leq 1$, the population variance in Equation 7.15 is larger than the implied population variance in Equation 7.14. Actually, they will be equal if $r_{X_2X_3}^2 = 0$, that is, if X_2 and X_3 are linearly uncorrelated. Moreover, when linearly uncorrelated $\sum_{i=1}^{n}(X_{2i}-\bar{X}_2)(X_{3i}-\bar{X}_3)=0$, then Equation 7.12 reduces to 7.13, meaning that including X_3 in the equation will not affect the estimation of β_2. While the population variances are the same, the estimated (sample) variances will still differ due to a reduction in the degrees of freedom.

7.3.1 Example

Consider the following model where we want to see how age affects the likelihood of being married. Are older people more likely to be married? Well, let's estimate the exact response of `married` to `age`,[1]

$$\texttt{married} = \beta_1 + \beta_2 \texttt{age} + \varepsilon \tag{7.16}$$

The estimation results from Gretl are

```
Model 1: OLS, using observations 1-935
Dependent variable: married

            coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const      0.540935     0.107608      5.027    5.98e-07 ***
  age        0.0106442    0.00323870    3.287    0.0011   ***

Mean dependent var   0.893048   S.D. dependent var   0.309217
Sum squared resid    88.28274   S.E. of regression   0.307608
R-squared            0.011445   Adjusted R-squared   0.010385
F(1, 933)            10.80160   P-value(F)           0.001052
Log-likelihood      -223.4066   Akaike criterion     450.8133
Schwarz criterion    460.4944   Hannan-Quinn         454.5047
```

$$\widehat{\texttt{married}} = \underset{(0.10761)}{0.540935} + \underset{(0.0032387)}{0.0106442}\,\texttt{age}$$

$$N = 935 \quad \bar{R}^2 = 0.0104 \quad F(1,933) = 10.802 \quad \hat{\sigma} = 0.30761$$

(standard errors in parentheses)

If the average `age` in the sample is 33 years of age, the predicted value for `married` is 89.2 ($\widehat{\texttt{married}} = 0.5409 + 0.0106 \times 33$). This means that if you are 33 years old, the probability that you are married is 89.2%. In addition, every year you get older, the probability that you are married increases by 0.011 or about 1.%. For some reason you think that only fools get married and then you decide to wrongly estimate the model

$$\texttt{married} = \beta_1 + \beta_2 \texttt{age} + \beta_3 \texttt{IQ} + \varepsilon \tag{7.17}$$

[1] Because `married` is actually a dummy variable this is a linear probability model, a type of model that we will see in detail in Chapter 9.

where the variable IQ is X_3 in Equation 7.11 and should not be in the model. The estimation results from Gretl are

```
Model 2: OLS, using observations 1-935
Dependent variable: married

             coefficient    std. error    t-ratio   p-value
  ----------------------------------------------------------
  const       0.563197      0.129804       4.339    1.59e-05 ***
  age         0.0106007     0.00324337     3.268    0.0011   ***
  IQ         -0.000205573   0.000669635   -0.3070   0.7589

Mean dependent var   0.893048   S.D. dependent var   0.309217
Sum squared resid    88.27381   S.E. of regression   0.307757
R-squared            0.011545   Adjusted R-squared   0.009424
F(2, 932)            5.442677   P-value(F)           0.004467
Log-likelihood      -223.3594   Akaike criterion     452.7187
Schwarz criterion    467.2404   Hannan-Quinn         458.2559
```

$$\widehat{\texttt{married}} = \underset{(0.12980)}{0.563197} + \underset{(0.0032434)}{0.0106007}\,\texttt{age} - \underset{(0.00066963)}{0.000205573}\,\texttt{IQ}$$

$$N = 935 \quad \bar{R}^2 = 0.0094 \quad F(2,932) = 5.4427 \quad \hat{\sigma} = 0.30776$$

(standard errors in parentheses)

Not surprisingly, the effect of IQ on married is not significant. This means that fools are not more likely to be married. However, the results do not necessarily support the conjecture that higher IQ is associated with married people either. Nevertheless, including IQ does not seems to help in the estimation of β_2. As we have seen theoretically, the estimate of the second equation is less efficient as can be appreciated from its larger standard error ($0.003243 > 0.003239$).

7.4 Testing a linear restriction

Testing linear restriction on the regression coefficients is sometimes very useful. Consider the following regression model,

$$\log \texttt{wage} = \beta_1 + \beta_2 \texttt{exper} + \beta_3 \texttt{educ} + \varepsilon \tag{7.18}$$

The regression output in Gretl is

```
Model 1: OLS, using observations 1-935
Dependent variable: logwage

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       5.50271      0.112037     49.12     8.13e-261 ***
  educ        0.0777820    0.00657687   11.83     3.62e-030 ***
  exper       0.0197768    0.00330251    5.988    3.02e-09  ***
```

```
Mean dependent var   6.779004   S.D. dependent var   0.421144
Sum squared resid    143.9786   S.E. of regression   0.393044
R-squared            0.130859   Adjusted R-squared   0.128994
F(2, 932)            70.16174   P-value(F)           4.13e-29
Log-likelihood      -452.0704   Akaike criterion     910.1407
Schwarz criterion    924.6624   Hannan-Quinn         915.6779
```

$$\widehat{\texttt{logwage}} = \underset{(0.11204)}{5.50271} + \underset{(0.0065769)}{0.0777820}\,\text{educ} + \underset{(0.0033025)}{0.0197768}\,\text{exper}$$

$$N = 935 \quad \bar{R}^2 = 0.1290 \quad F(2,932) = 70.162 \quad \hat{\sigma} = 0.39304$$

(standard errors in parentheses)

Let's say that we want to text whether the effect of a year on education on wages is the same as the effect of a year of experience of wages. That is, we want to text the following null hypothesis,

$$H_0 : \beta_2 = \beta_3 \tag{7.19}$$

While it may be tempting to just look and compare the regression estimates b_2 and b_3, this approach is not correct. Remember that b_2 and b_3 are just estimates and are not the unknown β_2 and β_3. The statistically correct approach is to run an auxiliary restricted regression where we force $b_2 = b_3$. Then, we have to compare if the regression fit with the *restricted* coefficients is significantly lower that the regression fit with the *unrestricted* (original) regression. To do this we calculate the residual sum of squares from the restricted model (RSS_R) and the residual sum of squares from the unrestricted model (RSS_U) and calculate the following F statistic:

$$F_{r,n-k} = \frac{(RSS_R - RSS_U)/r}{RSS_U/(n-k)} \tag{7.20}$$

where F is distributed with r and $n-k$ degrees of freedom. The number of restrictions r is equal to one in our example.

This is done automatically in Gretl. After you estimate the unrestricted regression model, in the regression output window you have to go to `Tests` → `Linear restrictions` and a new window will open. In the new window you have to type the command `b[educ] - b[exper] = 0` to obtain

```
Restriction:
 b[educ] - b[exper] = 0

Test statistic: F(1, 932) = 97.8892, with p-value = 5.14357e-022

Restricted estimates:

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       6.24122      0.0877816    71.10     0.0000    ***
  educ        0.0214837    0.00346501    6.200    8.46e-010 ***
  exper       0.0214837    0.00346501    6.200    8.46e-010 ***

  Standard error of the regression = 0.412948
```

The calculated F-statistics (that used Equation 7.20) is 97.8892 with an associated p-value that is below 0.05. This means that the fit in the two regression equations is significantly different and we reject the null hypothesis presented in Equation 7.19. We conclude that the effect of education and experience have a significantly different effect on wages.

If you want to test whether education had four times the effect on wages than experience, the null is

$$H_0 : \beta_2 = 4 \times \beta_3 \tag{7.21}$$

The command in Gretl is `b[educ] - 4*b[exper] = 0` to have

```
Restriction:
 b[educ] - 4*b[exper] = 0

Test statistic: F(1, 932) = 0.0126711, with p-value = 0.910399

Restricted estimates:

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       5.50597      0.108181      50.90    1.89e-271 ***
  educ        0.0778171    0.00656598    11.85    2.78e-030 ***
  exper       0.0194543    0.00164150    11.85    2.78e-030 ***

  Standard error of the regression = 0.392836
```

Notice that the F-statistics is fairly small and has a p-value that is now greater than 5%. We do not reject the null hypothesis and conclude that, on average, one year of education has four times the effect on wages than one year of experience.

Chapter 8
Heteroscedasticity

The fourth assumption in the estimation of the coefficients via ordinary least squares is the one of homoscedasticity. This means that the error terms u_i in the linear regression model have a constant variance across all observations i,

$$\sigma_{u_i}^2 = \sigma_u^2 \quad \text{for all } i. \tag{8.1}$$

When this assumption does not hold, and $\sigma_{u_i}^2$ changes across i we say we have an heteroscedasticity problem. This chapter discusses the problems associated with heteroscedastic errors, presents some tests for heteroscedasticity and points out some possible solutions.

8.1 Heteroscedasticity and its implications

What happens if the errors are heteroscedasticity? The good news is that under heteroscedastic errors, OLS is still unbiased. The bad news is that we will obtain the incorrect standard errors of the coefficients. This means that the t and the F tests that we discussed in earlier chapters are no longer valid. Figure 8.1 shows the regression equation `wage` = β_0 + β_1`educ` + u with heteroscedastic errors. The variance of u_i increases with higher values of `educ`.

8.2 Testing for heteroscedasticity

8.2.1 Breusch-Pagan test

Given the linear regression model

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_K + u \tag{8.2}$$

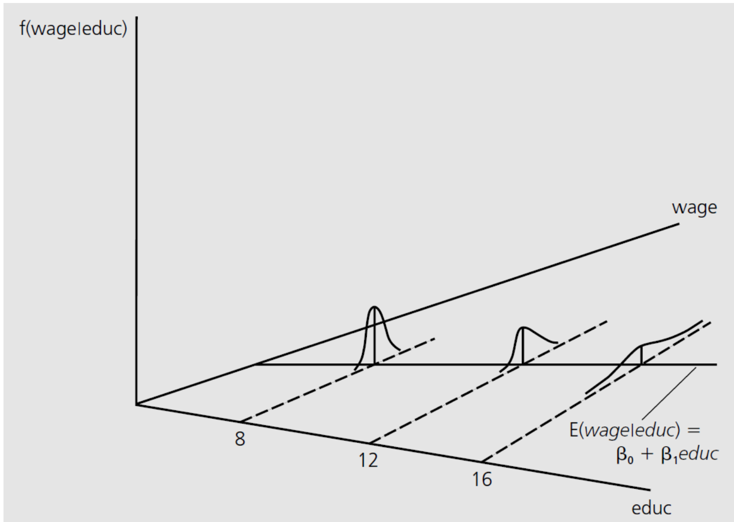

Fig. 8.1 `wage` = β_0 + β_1`educ` + u with heteroscedastic errors.

we know that OLS is unbiased and consistent if we assume $E[u|X_2, X_3, \ldots, X_K] = 0$. Let the null hypothesis that we have homoscedastic errors be

$$H_0 : Var[u|X_2, X_3, \ldots, X_K] = \sigma^2. \tag{8.3}$$

Because we are assuming that u has zero conditional expectation, $\mathrm{Var}[u|X_2, X_3, \ldots, X_K] = E[u^2|X_2, X_3, \ldots, X_K]$, and so the null hypothesis of homoscedasticity is equivalent to

$$H_0 : E[u^2|X_2, X_3, \ldots, X_K] = \sigma^2. \tag{8.4}$$

This shows that if we want to test for violation of the homoscedasticity assumption, we want to test whether $E[u^2|X_2, X_3, \ldots, X_K]$ is related to one or more of the independent variables. If H_0 is false, $E[u|X_2, X_3, \ldots, X_K]$ can be any function of the independent variables. A simple approach is to assume a linear function

$$u^2 = \delta_1 + \delta_2 X_1 + \delta_3 X_3 + \cdots + \delta_K X_K + \varepsilon, \tag{8.5}$$

where ε is an error term with mean zero given X_2, X_3, $\ldots$, X_K. The null hypothesis for homoscedasticity is:

$$H_0 : \delta_1 = \delta_2 = \delta_3 = \cdots = \delta_K = 0. \tag{8.6}$$

Under the null, it is reasonable to assume that ε is independent of $X_2, X_3, \ldots, X_K$. To be able to implement this test, we follow a two step procedure. In the first step we estimate Equation 5.14 via OLS. We estimate the residuals, square them and then estimate the following equation:

$$\hat{u}^2 = \delta_1 + \delta_2 X_1 + \delta_3 X_3 + \cdots + \delta_K X_K + error. \tag{8.7}$$

We can then easily compute the F statistic for the joint significance of all variables X_2, X_3, $\ldots$, X_K. Using OLS residuals in place of the errors does not affect the large sample distribution of the F statistic. An additional LM statistic to test for heteroscedasticity can be constructed based on the $R^2_{\hat{u}^2}$ obtained from Equation 8.7:

$$LM = n \cdot R^2_{\mathrm{u}^2}. \tag{8.8}$$

Under the null hypothesis, LM is distributed asymptotically as χ^2_{K-1}. This LM version of the test is called the Breusch-Pagan test for heteroscedasticity.

8.2.2 Breusch-Pagan test in Gretl

As an example, consider once again our wage equation

$$\texttt{wage} = \beta_1 + \beta_2 \texttt{educ} + u \tag{8.9}$$

Once we estimated the model in Gretl

$$\widehat{\text{wage}} = \underset{(77.715)}{146.952} + \underset{(5.6950)}{60.2143}\,\texttt{educ}$$

$$N = 935 \quad \bar{R}^2 = 0.1060 \quad F(1,933) = 111.79 \quad \hat{\sigma} = 382.32$$

(standard errors in parentheses)

In the regression output window, go to `Tests` → `Heteroskedasticity` → `Breusch-Pagan` to obtain

```
Breusch-Pagan test for heteroskedasticity
OLS, using observations 1-935
Dependent variable: scaled uhat^2

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       -0.885844    0.450097     -1.968    0.0494   **
  educ         0.140019    0.0329833     4.245    2.40e-05 ***

  Explained sum of squares = 88.3581

Test statistic: LM = 44.179066,
with p-value = P(Chi-square(1) > 44.179066) = 0.000000
```

Notice how Gretl reports the auxiliary regression presented in Equation 8.7 and the LM statistic from Equation 8.8. The large LM statistic associated with a small p-value (below 0.05 or 5%) indicates that we reject the null hypothesis of homoscedasticity. Hence, we have heteroscedaticity in the model of Equation 8.9.

8.2.3 White test

White (1980) proposed a test for heteroscedasticity that that adds the squares and cross products of all the independent variables to Equation 8.2. In a model with only three independent variables, the White test is based on the estimation of:

$$\begin{aligned}\hat{u}^2 = &\ \delta_1 + \delta_2 X_2 + \delta_3 X_3 + \delta_4 X_4 + \delta_5 X_2^2 + \delta_6 X_3^2 + \delta_7 X_4^2 \\ &\ \delta_8 X_2 \cdot X_3 + \delta_9 X_2 \cdot X_4 + \delta_{10} X_3 \cdot X_4 + error.\end{aligned} \tag{8.10}$$

Compared with the Breusch-Pagan test (see Equation 8.7), Equation 8.10 has more regressors. The White test for heteroscedasticity is based on the *LM* statistic for testing that all the δ_j in Equation 8.10 are zero, except for the intercept.

8.2.4 White test in Gretl

We not use Gretl to test for heteroscedasticity in Equation 8.9 using the White test. In the regression output window, go to `Tests` → `Heteroskedasticity` → `White's test` to obtain

```
White's test for heteroskedasticity
OLS, using observations 1-935
Dependent variable: uhat^2

             coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const      -126650        435765       -0.2906     0.7714
  educ         20049.7       63065.6      0.3179     0.7506
  sq_educ         13.2563     2234.63     0.005932   0.9953

  Unadjusted R-squared = 0.018950

Test statistic: TR^2 = 17.717812,
with p-value = P(Chi-square(2) > 17.717812) = 0.000142
```

Consistent with the Breusch-Pagan test, here the White test has a large *LM* statistic (labeled `TR^2` following $LM = n \cdot R^2_{u^2}$ as in Equation 8.8) associated with a small p-value (smaller than 5%). Hence, we reject the null of homoscedasticity and conclude that our model is heteroscedastic.

8.3 What to do with heteroscedasticity?

There is a number of possible solutions when heteroscedastic errors are found. This section proposes three ways to solve the heteroscedasticity problem. First, a simple transformation of the variables; second, the use of weighted least squares; and third, the use of heteroscedasticity-robust standard errors.

8.3.1 Simple transformation of the variables

An easy way to obtain homoscedastic errors is to come up with a simple transformation of the variables. Let's revisit the estimation of Equation 8.9, but this time with a simple logarithm transformation of wages,

$$\log \texttt{wage} = \beta_1 + \beta_2 \texttt{educ} + u \tag{8.11}$$

The Gretl regression output is

$$\widehat{\texttt{logwage}} = \underset{(0.081374)}{5.97306} + \underset{(0.0059631)}{0.0598392}\,\texttt{educ}$$
$$N = 935 \quad \bar{R}^2 = 0.0964 \quad F(1,933) = 100.70 \quad \hat{\sigma} = 0.40032$$
(standard errors in parentheses)

Now, if we want to test for the existence of heteroscedasticity we go to `Tests` → `Heteroskedasticity` → `Breusch-Pagan` to obtain

```
Breusch-Pagan test for heteroskedasticity
OLS, using observations 1-935
Dependent variable: scaled uhat^2

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const       0.689778     0.329663     2.092     0.0367  **
  educ        0.0230332    0.0241578    0.9534    0.3406

  Explained sum of squares = 2.391

Test statistic: LM = 1.195499,
with p-value = P(Chi-square(1) > 1.195499) = 0.274223
```

Notice that the p-value associated with this test is above 0.05. Hence, we fail to reject the null of homoscedasticity. Compare this homoscedasticity results with the heteroscedastic errors found earlier when the variable `wage` was not in logs.

8.3.2 Weighted Least Squares

We want to estimate the following regression model

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_K X_K + u, \tag{8.12}$$

but the errors u are heteroscedastic. When one is willing to assume that the heteroscedasticity appears as some function of X_2, X_3, ..., X_K, one can use Weighted Least Squares (WLS) to obtain homoscedastic errors. Let's say that the variance of u can be approximated using

$$u^2 = \sigma^2 \exp(\delta_1 + \delta_2 X_2 + \delta_3 X_3 + \cdots + \delta_K X_K)\eta, \tag{8.13}$$

where η is a random variable with mean equal to unity. If we assume that η is independent from X_2, X_3, …, X_K we have

$$\log(u^2) = \alpha_1 + \delta_2 X_2 + \delta_3 X_3 + \cdots + \delta_K X_K + \varepsilon. \tag{8.14}$$

To be able to implement this procedure, we replace the unobserved u with the OLS estimated residuals $\hat{u}$ to estimate:

$$\log(\hat{u}^2) = \alpha_1 + \delta_2 X_2 + \delta_3 X_3 + \cdots + \delta_K X_K + \varepsilon. \tag{8.15}$$

Finally, once Equation 8.15 is estimated, we obtain the fitted values and calculate the exponent to obtain

$$\hat{h}_i = \exp(\widehat{\log(\hat{u}^2)}). \tag{8.16}$$

We can use this $\hat{h}_i$ as a weight in a Weighted Least Squares regression to solve the heteroscedasticity problem. That is, we estimate the following weighted equation

$$\frac{Y}{\hat{h}} = \beta_1 \frac{1}{\hat{h}} + \beta_2 \frac{X_2}{\hat{h}} + \beta_3 \frac{X_3}{\hat{h}} + \cdots + \beta_K \frac{X_K}{\hat{h}} + \frac{u}{\hat{h}}. \tag{8.17}$$

Notice that Equation 8.17 is just Equation 8.12 divided by the weight $\hat{h}_i$. The new error term $u/\hat{h}$ should be homoscedastic.

8.3.3 Weighted Least Squares in Gretl

Consider the following model

$$\mathtt{sav} = \beta_1 + \beta_2 \mathtt{inc} + u \tag{8.18}$$

where `sav` is savings and `inc` is income. The Gretl output is

```
Model 1: OLS, using observations 1-100
Dependent variable: sav

             coefficient   std. error    t-ratio   p-value
  --------------------------------------------------------
  const      124.842       655.393       0.1905    0.8493
  inc          0.146628      0.0575488   2.548     0.0124   **

Mean dependent var   1582.510   S.D. dependent var   3284.902
Sum squared resid    1.00e+09   S.E. of regression   3197.415
R-squared            0.062127   Adjusted R-squared   0.052557
F(1, 98)             6.491778   P-value(F)           0.012391
Log-likelihood      -947.8935   Akaike criterion     1899.787
Schwarz criterion    1904.997   Hannan-Quinn         1901.896
```

and the Breusch-Pagan test for heteroscedasticity yields

```
Breusch-Pagan test for heteroskedasticity
OLS, using observations 1-100
Dependent variable: scaled uhat^2

             coefficient   std. error    t-ratio   p-value
  --------------------------------------------------------
  const      0.0457266     1.14381       0.03998   0.9682
  inc        9.59914e-05   0.000100436   0.9557    0.3416

  Explained sum of squares = 28.444

Test statistic: LM = 14.221987,
with p-value = P(Chi-square(1) > 14.221987) = 0.000162
```

That is, we have heteroscedastic errors.

To estimate the WLS regression

$$\frac{\texttt{sav}}{\hat{h}} = \beta_1 \frac{1}{\hat{h}} + \beta_2 \frac{\texttt{inc}}{\hat{h}} + \frac{u}{\hat{h}}, \tag{8.19}$$

in the Gretl main window we have to go to `Model` → `Other linear models` → `Heteroskedasticity corrected` to get the following computer output

```
Model 2: Heteroskedasticity-corrected, using observations 1-100
Dependent variable: sav

             coefficient   std. error    t-ratio   p-value
  --------------------------------------------------------
  const      -233.130      460.844       -0.5059   0.6141
  inc           0.185993     0.0616965    3.015    0.0033  ***

Statistics based on the weighted data:

Sum squared resid    1043.864  S.E. of regression   3.263689
R-squared            0.084866  Adjusted R-squared   0.075527
F(1, 98)             9.088089  P-value(F)           0.003276
Log-likelihood      -259.1695  Akaike criterion     522.3391
Schwarz criterion    527.5494  Hannan-Quinn         524.4478

Statistics based on the original data:

Mean dependent var   1582.510  S.D. dependent var   3284.902
Sum squared resid    1.01e+09  S.E. of regression   3205.216
```

$$\widehat{\texttt{sav}} = \underset{(460.84)}{-233.130} + \underset{(0.061697)}{0.185993}\,\texttt{inc}$$

$$N = 100 \quad \bar{R}^2 = 0.0755 \quad F(1,98) = 9.0881 \quad \hat{\sigma} = 3.2637$$

(standard errors in parentheses)

The estimates of the standard errors can now be used for inferences. The statistically significant coefficient on `inc` indicates that the marginal propensity to save out of your income is 0.18. Of every additional dollar that you make, you will save 18 cents.

8.3.4 White's heteroscedasticity-consistent standard errors

Even under the presence of heteroscedastic errors, at least in large samples a consistent estimator of the variances of the coefficients can be obtained via White's heteroscedasticity-consistent standard errors. This procedure leaves the OLS coefficients unaffected. For the estimation of Equation 8.18 you just have to make sure to select the option `Robust standard errors` in the Gretl "specify model" window when you estimate the model via OLS

```
Model 3: OLS, using observations 1-100
Dependent variable: sav
Heteroskedasticity-robust standard errors, variant HC1

             coefficient   std. error    t-ratio   p-value
  ---------------------------------------------------------
  const      124.842       528.219       0.2363    0.8137
  inc          0.146628      0.0613441   2.390     0.0187  **

Mean dependent var   1582.510   S.D. dependent var   3284.902
Sum squared resid    1.00e+09   S.E. of regression   3197.415
R-squared            0.062127   Adjusted R-squared   0.052557
F(1, 98)             5.713342   P-value(F)           0.018748
Log-likelihood      -947.8935   Akaike criterion     1899.787
Schwarz criterion    1904.997   Hannan-Quinn         1901.896
```

Notice that the constant and slope coefficients are the same as before. However, the estimated standard errors are different.

Chapter 9
Binary Choice Models

Some time we are interested in analyzing *binary response* or *qualitative response variables* that have outcomes Y equal to 1 when the even occurs and equal to 0 when the event does not occur. Some example include going to college, getting married, buying a house, or getting a job. All these cases involve a yes/no answer. How is this yes/no answer affected by other variables? That is the subject matter of this chapter.

9.1 The linear probability model

9.1.1 The model

The simplest binary choice model is the *linear probability model*, where as its name suggests, the probability of the event occurring, p, is assumed to be a linear function of a set of explanatory variable. If we only have one variable the model is

$$p_i = p(Y_i = 1) = \beta_1 + \beta_2 X_i. \tag{9.1}$$

The response variable Y_i can be written as the summation of its deterministic and its random component,

$$Y_i = E(Y_i|X_i) + u_i. \tag{9.2}$$

It is simple to compute $E(Y_i|X_i)$, the expected value of Y_i given X_i, because Y takes only two values. It is 1 with probability p_i and 0 with probability $1 - p_i$,

$$E(Y_i|X_i) = 1 \times p_i + 0 \times (1 - p_i) = p_i = \beta_1 + \beta_2 X_i. \tag{9.3}$$

This means that we can write the model as

$$Y_i = \beta_1 + \beta_2 X_i + u_i, \tag{9.4}$$

that is just the same model we have been estimating in previous chapters. The big difference is that Y_i takes only the values 0 and 1.

9.1.2 The linear probability model in Gretl

Let's estimate the following model

$$\texttt{inlf}_i = \beta_1 + \beta_2 \texttt{educ}_i + \beta_3 \texttt{faminc}_i + u_i, \tag{9.5}$$

where `inlf` is equal to one if individual i is in the labor force, zero otherwise, `educ` is the number of years of education and `faminc` is the family income. The regression command in Gretl is the same as before.

```
Model 1: OLS, using observations 1-753
Dependent variable: inlf

             coefficient   std. error    t-ratio   p-value
  --------------------------------------------------------
  const      0.0689887     0.0973736     0.7085    0.4789
  educ       0.0379040     0.00835610    4.536     6.67e-06 ***
  faminc     1.45940e-06   1.56306e-06   0.9337    0.3508

Mean dependent var   0.568393   S.D. dependent var   0.495630
Sum squared resid    178.0367   S.E. of regression   0.487219
R-squared            0.036221   Adjusted R-squared   0.033651
F(2, 750)            14.09349   P-value(F)           9.81e-07
Log-likelihood      -525.5192   Akaike criterion     1057.038
Schwarz criterion    1070.911   Hannan-Quinn         1062.383
```

$$\widehat{\texttt{inlf}} = \underset{(0.097374)}{0.0689887} + \underset{(0.0083561)}{0.0379040}\,\texttt{educ} + \underset{(1.5631\text{e–}006)}{1.45940\text{e–}006}\,\texttt{faminc}$$

$$N = 753 \quad \bar{R}^2 = 0.0337 \quad F(2,750) = 14.093 \quad \hat{\sigma} = 0.48722$$

(standard errors in parentheses)

For example, the coefficient on `educ` indicates that every additional years of education increases the probability of being in the labor force by about 4%. This information is graphed in Figure 9.1.

There are two main problems with a linear probability model such as the one presented in Equation 9.4.

1. The model will predict unrealistic probabilities, beyond 1 and below 0 (see Figure 9.1).
2. Because Y_i only takes the values of 0 and 1, the error term u will be far from following a normal distribution.

The solution is to transform the linear probability model. Two common transformations are the *logit* and the *probit*.

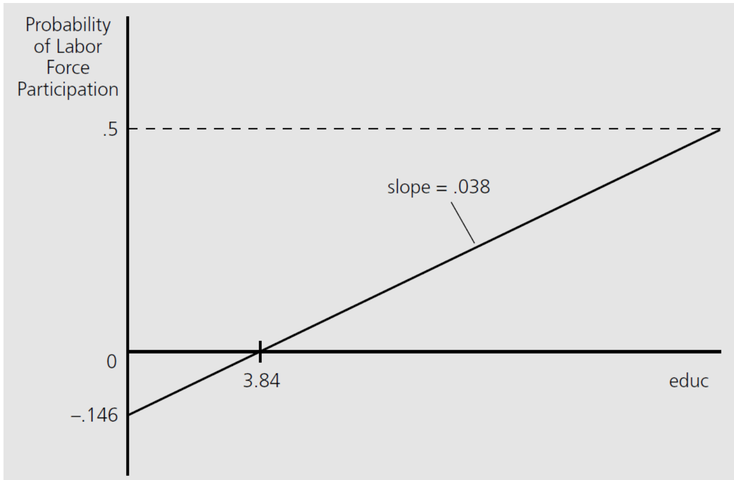

Fig. 9.1 $\texttt{inlf}_i = \beta_1 + \beta_2 \texttt{educ}_i + \beta_3 \texttt{faminc}_i + u_i$

9.2 Logit analysis

9.2.1 The logit transformation

Let Z_i be,

$$Z_i = \beta_1 + \beta_2 X_i \tag{9.6}$$

The logit model hypothesizes that the probability of occurrence of the event $Y = 1$ is determined by the function

$$p_i = F(Z_i) = \frac{1}{1+e^{-Z}} \tag{9.7}$$

where

$$\frac{\partial p}{\partial Z} = f(Z) = \frac{e^{-Z}}{(1+e^{-Z})^2} \tag{9.8}$$

and

$$\frac{\partial p}{\partial X} = \frac{\partial p}{\partial Z}\frac{\partial Z}{\partial X} = f(Z)\beta_2 \tag{9.9}$$

This means that the marginal effect of variable X on the probability of $Y = 1$ is $f(Z)\beta_2$, where $f(Z)$ needs to be evaluated on some specific value of X, let's say, the mean of X.

9.2.2 Logit regression in Gretl

Fortunately, all these calculations are done automatically by Gretl. If we want to obtain the logit estimates of Equation 9.5 in the main Gretl window we have to go to `Model` → `Nonlinear models` → `Logit` → `Binary...` and select the option "Show p-values" to obtain

```
Convergence achieved after 4 iterations

Model 2: Logit, using observations 1-753
Dependent variable: inlf

             coefficient    std. error       z      p-value
  ---------------------------------------------------------
  const      -1.85287       0.428444      -4.325    1.53e-05 ***
  educ        0.161773      0.0367856      4.398    1.09e-05 ***
  faminc      6.58050e-06   6.83134e-06    0.9633   0.3354

Mean dependent var   0.568393   S.D. dependent var   0.244933
McFadden R-squared   0.027185   Adjusted R-squared   0.021359
Log-likelihood      -500.8762   Akaike criterion     1007.752
Schwarz criterion    1021.625   Hannan-Quinn         1013.097

Number of cases 'correctly predicted' = 449 (59.6%)
f(beta'x) at mean of independent vars = 0.245
Likelihood ratio test: Chi-square(2) = 27.9939 [0.0000]

           Predicted
             0     1
  Actual 0   69   256
         1   48   380
```

Gretl actually estimates this model using an estimation technique called Maximum Likelihood Estimation, that is why the computer iterates before giving the estimates. The output is very similar as the one obtained in previous chapters. The effect of `educ` on `inlf` is statistically significant. However, the key difference in this output is that the coefficients are not interpreted as the marginal effects. Recall that the marginal effects are calculated using Equation 9.9. To make Gretl obtain this marginal effects you need to reestimate the model and select the option "Show slopes at mean" to obtain

```
             coefficient    std. error       z         slope
  -------------------------------------------------------------
  const      -1.85287       0.428444      -4.325
  educ        0.161773      0.0367856      4.398    0.0396234
  faminc      6.58050e-06   6.83134e-06    0.9633   1.61178e-06
```

The marginal effect of `educ` on `inlf` is actually 0.0396. An additional year of education will increase the probability that you are in the labor force by about 4%.

9.3 Probit analysis

9.3.1 The probit transformation

The probit model is similar in spirit as the logit model. Let Z_i be,

$$Z_i = \beta_1 + \beta_2 X_i \tag{9.10}$$

The probit model hypothesizes that the probability of occurrence of the event $Y = 1$ is determined by the function

$$p_i = F(Z_i) \tag{9.11}$$

where $F(\cdot)$ is actually the cumulative standardized normal distribution. Then,

$$\frac{\partial p}{\partial Z} = f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \tag{9.12}$$

is just the derivative of $F(\cdot)$. As in the logit case,

$$\frac{\partial p}{\partial X} = \frac{\partial p}{\partial Z}\frac{\partial Z}{\partial X} = f(Z)\beta_2 \tag{9.13}$$

Again, this means that the marginal effect of variable X on the probability of $Y = 1$ is $f(Z)\beta_2$, where $f(Z)$ needs to be evaluated on some specific value of X, let's say, the mean of X.

9.3.2 Probit regression in Gretl

If we want to obtain the probit estimates of Equation 9.5 in the main Gretl window we have to go to `Model` → `Nonlinear models` → `Probit` → `Binary...` and select the option "Show p-values" to obtain

```
Convergence achieved after 5 iterations

Model 3: Probit, using observations 1-753
Dependent variable: inlf

             coefficient    std. error       z       p-value
  ---------------------------------------------------------
  const      -1.14768       0.261470      -4.389    1.14e-05 ***
  educ        0.100666      0.0224296      4.488    7.19e-06 ***
  faminc      3.84752e-06   4.09647e-06    0.9392   0.3476

Mean dependent var   0.568393   S.D. dependent var   0.392673
McFadden R-squared   0.027216   Adjusted R-squared   0.021389
Log-likelihood      -500.8606   Akaike criterion     1007.721
Schwarz criterion    1021.593   Hannan-Quinn         1013.065
```

```
Number of cases 'correctly predicted' = 449 (59.6%)
f(beta'x) at mean of independent vars = 0.393
Likelihood ratio test: Chi-square(2) = 28.0252 [0.0000]

           Predicted
             0     1
  Actual 0   69   256
         1   48   380
```

Once again, the effect of `educ` on `inlf` is statistically significant. To make Gretl obtain this marginal effects using Equation 9.13 you need to reestimate the model and select the option "Show slopes at mean" to obtain

```
             coefficient    std. error       z         slope
    -------------------------------------------------------------
    const      -1.14768       0.261470      -4.389
    educ        0.100666      0.0224296      4.488    0.0395289
    faminc      3.84752e-06   4.09647e-06    0.9392   1.51081e-06
```

The marginal effect of `educ` on `inlf` is 0.0395. We obtain almost the same results as before. The probit model predicts that an additional year of education will increase the probability that you are in the labor force by about 4%.

Chapter 10
Time Series

10.1 Time Series Data

The main difference between time series data and cross-sectional data is the temporal ordering. To emphasize the proper ordering of the observations, Table 10.1 presents a partial listing of the data on U.S. inflation and unemployment rates from 1948 through 2003. Unlike cross-sectional data, in time series the temporal order in which the observations appear in the data set is very important. In terms of notation, we use the subscript t to denote time and we use it instead of the previous subscript i, i.e., X_t.

Table 10.1 U.S. Inflation and Unemployment Rates, 1965-2011

Year	Inflation	Unemployment
1948	8.1	3.8
1949	-1.2	5.9
1950	1.3	5.3
1951	7.9	3.3
⋮	⋮	⋮
2000	3.4	4.0
2001	2.8	4.7
2002	1.6	5.8
2003	2.3	6.0

A second key difference between time series and cross-sectional data is that in the latter we assume that the sample was randomly drawn from the population. While in time series the variables are also considered random, a variable indexed by time is called a *stochastic process* or a *time series process*. When we collect a time series data set we are one possible outcome or realization of the stochastic process. We can only see a single realization because we cannot go back in time and start the process again.

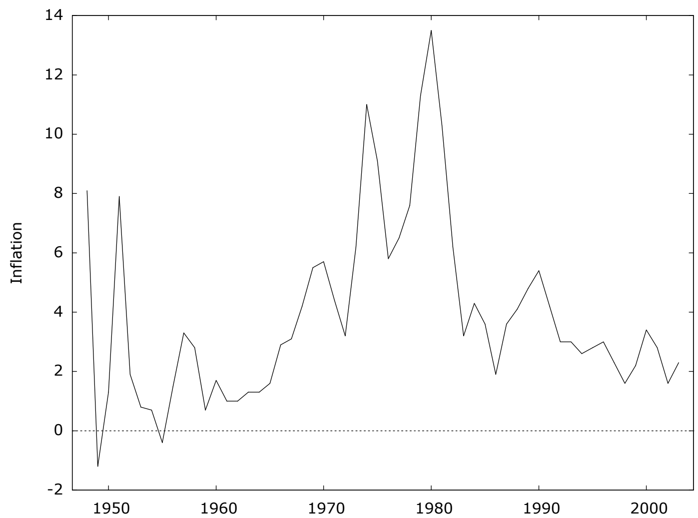

Fig. 10.1 Inflation, 1948-2003.

Graphing the data is particularly important to visualize the dynamics of the variables. Figure 10.1 presents the time series graph of inflation from 1948 through 2003. One can easily identify the periods of high inflation late in the seventies and early eighties. To obtain this graph in Gretl, go to `View` → `Graph specified vars` → `Time series plot` and then select the variables you want to plot against time.

10.2 Time Series Regression Models

10.2.1 Static Models

The simplest *static model* has the form

$$Y_t = \beta_1 + \beta_2 X_t + u_t, \quad t = 1, 2, 3, \ldots, n. \tag{10.1}$$

We call this a static model because we are only modeling a contemporaneous relationship between X_t and Y_t. That is, when a change in X at time t is believed to have an immediate effect on Y: $\Delta Y_t = \beta_2 \Delta X_t$. One example is the static Phillips curve given by:

$$\texttt{inflation}_t = \beta_1 + \beta_2 \texttt{unemployment}_t + u_t. \tag{10.2}$$

where `inflation` is the annual inflation rate, and `unemployment` is the unemployment rate. Estimation in Gretl via OLS is follows the same steps as in the previous chapters. The output for the estimation of Equation 10.1 is:

```
Model 1: OLS, using observations 1948-2003 (T = 56)
Dependent variable: inflation

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const         1.05357      1.54796     0.6806    0.4990
  unemployment  0.502378     0.265562    1.892     0.0639 *

Mean dependent var   3.883929   S.D. dependent var   3.040381
Sum squared resid    476.8157   S.E. of regression   2.971518
R-squared            0.062154   Adjusted R-squared   0.044786
F(1, 54)             3.578726   P-value(F)           0.063892
Log-likelihood      -139.4304   Akaike criterion     282.8607
Schwarz criterion    286.9114   Hannan-Quinn         284.4311
rho                  0.572055   Durbin-Watson        0.801482
```

$$\widehat{\texttt{inflation}} = \underset{(1.5480)}{1.05357} + \underset{(0.26556)}{0.502378}\,\texttt{unemployment}$$

$$T = 56 \quad \bar{R}^2 = 0.0448 \quad F(1,54) = 3.5787 \quad \hat{\sigma} = 2.9715$$

(standard errors in parentheses)

The estimation results indicate that a one point increase in the unemployment rate is linked with a 0.5 increase in the inflation rate. Of course more variables can be included in the model. Notice that we can use this model to predict `inflation` given that we know the values for `unemployment` by simply plugging values for unemployment in the estimated equation. If we do this for the actual unemployment values for 1948-2003 period and graph them, we obtain the fitted values graph. Figure 10.2 plots the actual and the fitted values for inflation.

10.2.2 Finite Distributed Lag Models

The simplest dynamic model is the *finite distributed lag* (FDL) model, where we allow one or more variables to to affect Y_t with a lag. Consider the following example:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + u_t, \tag{10.3}$$

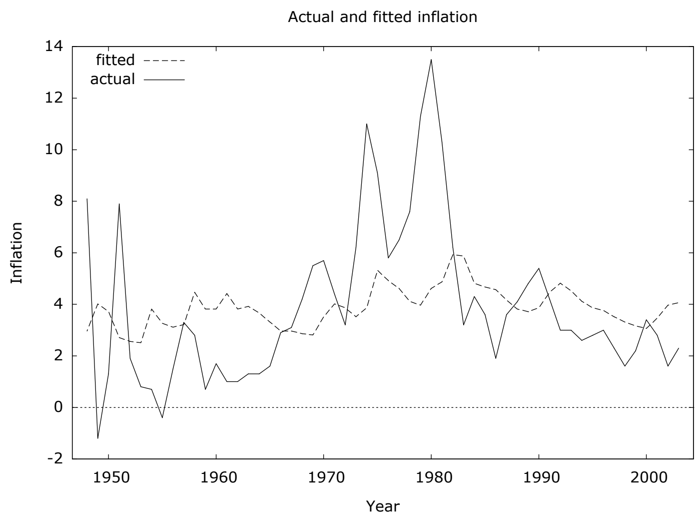

Fig. 10.2 Inflation, 1948-2003. Actual and fitted based on an static model.

where the FDL is of order two. Let's say that we are interested in the effect on Y of a permanent increase in X. Before time t, X equals to a constant c. At time t, X increases permanently to $c+1$. That is, $X_s = c$ for $s < t$ and $X_s = c+1$ for $s \geq t$. Setting the errors to be equal to zero we have:

$$\begin{aligned} Y_{t-1} &= \beta_1 + \beta_2 c + \beta_3 c + \beta_4 c \\ Y_t &= \beta_1 + \beta_2 (c+1) + \beta_3 c + \beta_4 c \\ Y_{t+1} &= \beta_1 + \beta_2 (c+1) + \beta_3 (c+1) + \beta_4 c \\ Y_{t+2} &= \beta_1 + \beta_2 (c+1) + \beta_3 (c+1) + \beta_4 (c+1) \end{aligned} \tag{10.4}$$

and so on. The contemporaneous effect of X on Y is called the *impact multiplier* and in this case this one is given by β_2. However, over time the marginal effect of X on Y is larger. We say that the *long-run multiplier* is the long-run change Y given a permanent increase in X. This one is given by $\beta_2 + \beta_3 + \beta_4$.

Consider the following example in Gretl:

$$\texttt{inf}_t = \beta_1 + \beta_2 \texttt{unem}_t + \beta_3 \texttt{unem}_{t-1} + \beta_4 \texttt{unem}_{t-2} + u_t, \tag{10.5}$$

To estimate this model in Gretl we first need to create the lagged values of `unem`. To do this we have to go to select `unem` and then go to `Add` → `Lags of selected variables` and select the number of lags. An alternative approach is to just include the lags when estimating the model via OLS. That is, when specifying the model in Gretl (`Model` → `Ordinary Least Squares`) there is an icon that allows you to select the lags. Just select two lags for `unem` to obtain:

```
Model 2: OLS, using observations 1950-2003 (T = 54)
Dependent variable: inf

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       -0.124609     1.68922     -0.07377   0.9415
  unem         0.903211     0.402071     2.246     0.0291  **
  unem_1      -0.856337     0.525700    -1.629     0.1096
  unem_2       0.668123     0.386722     1.728     0.0902  *

Mean dependent var   3.900000   S.D. dependent var   2.961323
Sum squared resid    395.2340   S.E. of regression   2.811526
R-squared            0.149632   Adjusted R-squared   0.098610
F(3, 50)             2.932693   P-value(F)           0.042366
Log-likelihood      -130.3660   Akaike criterion     268.7320
Schwarz criterion    276.6880   Hannan-Quinn         271.8003
rho                  0.661217   Durbin-Watson        0.676987
```

$$\widehat{\texttt{inf}} = \underset{(1.6892)}{-0.124609} + \underset{(0.40207)}{0.903211}\,\texttt{unem} - \underset{(0.52570)}{0.856337}\,\texttt{unem\_1} + \underset{(0.38672)}{0.668123}\,\texttt{unem\_2}$$

$$T = 54 \quad \bar{R}^2 = 0.0986 \quad F(3,50) = 2.9327 \quad \hat{\sigma} = 2.8115$$

(standard errors in parentheses)

A permanent increase in unemployment leads to a contemporaneous increase in inflation of 0.903 (impact multiplier). However, in the long-run the same increase in unemployment leads to a permanent effect on inflation of 0.903 -0.856 + 0.668 = 0.715 (long-run multiplier).

10.2.3 Autoregressive Model

An *autoregresive model* is a simple model where the current values of a variable are related to its past values. The first-order autoregressive model is given by:

$$Y_t = \phi Y_{t-1} + u_t. \tag{10.6}$$

This one is usually denoted by $AR(1)$. A more general model is the pth autoregressive model or $AR(p)$ given by:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \cdots + \phi_p Y_{t-p} + u_t \tag{10.7}$$

where there are p lags of the variable Y explaining its current value. The estimation of an $AR(p)$ model in Gretl is simple; go to `Model` → `Time series` → `ARIMA` and then select the dependent variable and the *AR* order. Make sure that the *MA* order is zero. For the example above, consider estimating the following model:

$$\texttt{inf}_t = \phi_1 \texttt{inf}_{t-1} + \phi_2 \texttt{inf}_{t-2} + u_t \tag{10.8}$$

The output in Gretl is:

```
Function evaluations: 17
Evaluations of gradient: 8

Model 4: ARMA, using observations 1948-2003 (T = 56)
Estimated using Kalman filter (exact ML)
Dependent variable: inf
Standard errors based on Hessian

             coefficient   std. error     z      p-value
  -------------------------------------------------------
  const       4.02526      0.791433     5.086   3.66e-07 ***
  phi_1       0.815712     0.148739     5.484   4.15e-08 ***
  phi_2      -0.175228     0.152459    -1.149   0.2504

Mean dependent var   3.883929   S.D. dependent var   3.040381
Mean of innovations -0.054104   S.D. of innovations  2.171763
Log-likelihood      -123.2506   Akaike criterion     254.5012
Schwarz criterion    262.6026   Hannan-Quinn         257.6421

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1           2.3276    -0.5378     2.3889    -0.0361
    Root  2           2.3276     0.5378     2.3889     0.0361
  -----------------------------------------------------------
```

That yields the following estimated equation:

$$\widehat{\texttt{inf}}_t = 4.025 + 0.8157\texttt{inf}_{t-1} - 0.1752\texttt{inf}_{t-2}. \tag{10.9}$$

where we can see that higher inflation last period has a positive effect on inflation this period. We can use this model to predict the path of `inf` based on its previous values. It Gretl the command to obtain this `Graphs` → `Fitted, actual plot` → `Against time`. The resulting graph is shown in Figure 10.3.

10.2.4 Moving-Average Models

The *moving-average* models express an observed series as a function of the current and lagged unobserved shocks. The simplest moving-average model is the moving-average of order one, or $MA(1)$:

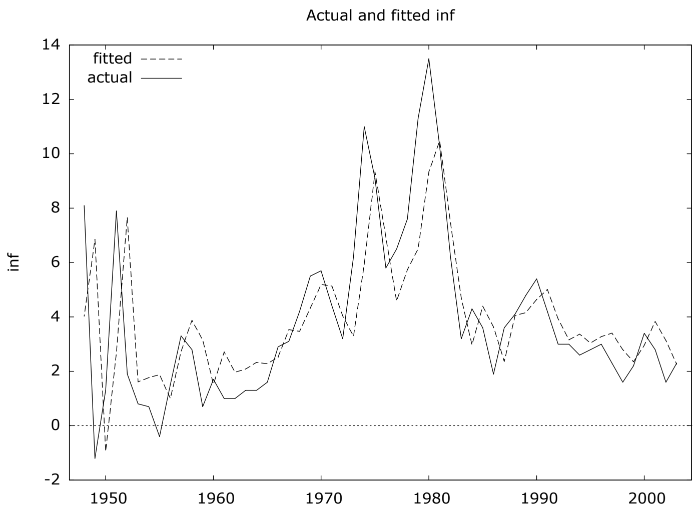

Fig. 10.3 Inflation, 1948-2003. Actual and fitted based on an $AR(2)$ model.

$$Y_t = \theta u_{t-1} + u_t \tag{10.10}$$

A more general moving-average of order q is be written as:

$$Y_t = \theta_1 u_{t-1} + \theta_2 u_{t-2} + \theta_3 u_{t-3} + \cdots + \theta_q u_{t-q} + u_t \tag{10.11}$$

For the example above:

$$\texttt{inf}_t = \theta_1 u_{t-1} + \theta_2 u_{t-2} + u_t \tag{10.12}$$

the output in Gretl is:

```
Function evaluations: 51
Evaluations of gradient: 19

Model 6: ARMA, using observations 1948-2003 (T = 56)
Estimated using Kalman filter (exact ML)
Dependent variable: inf
Standard errors based on Hessian

             coefficient   std. error     z      p-value
```

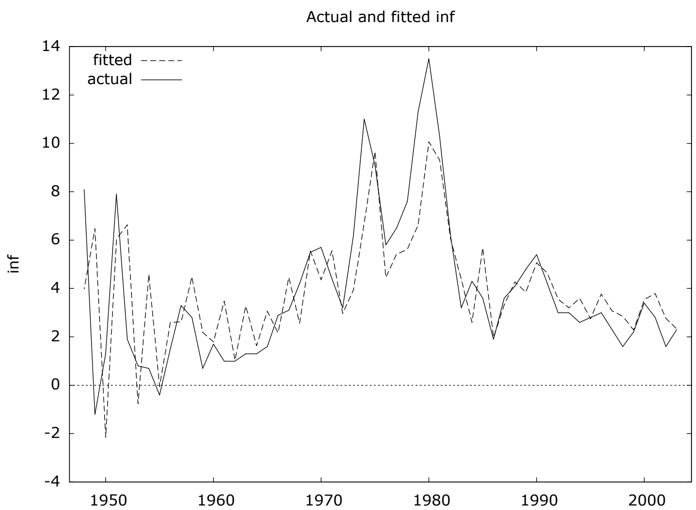

Fig. 10.4 Inflation, 1948-2003. Actual and fitted based on an $MA(2)$ model.

```
  -------------------------------------------------------
  const        3.98267       0.615240    6.473   9.58e-011 ***
  theta_1      1.18549       0.130300    9.098   9.19e-020 ***
  theta_2      0.267922      0.127516    2.101   0.0356    **

Mean dependent var   3.883929   S.D. dependent var   3.040381
Mean of innovations -0.041523   S.D. of innovations  1.899580
Log-likelihood      -116.5030   Akaike criterion     241.0061
Schwarz criterion    249.1075   Hannan-Quinn         244.1470

                      Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  MA
    Root  1          -1.1343     0.0000     1.1343     0.5000
    Root  2          -3.2904     0.0000     3.2904     0.5000
  -----------------------------------------------------------
```

and the actual and fitted values are presented in Figure 10.4.

10.2.5 Autoregressive Moving Average Models

One can easily combine an $AR(1)$ model and an $MA(1)$ models to obtain an autoregressive moving-average model $ARMA(1,1)$:

$$Y_t = \phi Y_{t-1} + \theta u_{t-1} + u_t \tag{10.13}$$

or a more general $ARMA(p,q)$ model:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} + u_t \tag{10.14}$$

The output in Gretl for a $ARMA(2,2)$ for inflation is:

```
Function evaluations: 61
Evaluations of gradient: 20

Model 5: ARMA, using observations 1948-2003 (T = 56)
Estimated using Kalman filter (exact ML)
Dependent variable: inf
Standard errors based on Hessian

             coefficient   std. error      z      p-value
  -------------------------------------------------------
  const       3.94843      1.05291       3.750    0.0002  ***
  phi_1       0.828806     0.236639      3.502    0.0005  ***
  phi_2       0.0226838    0.173277      0.1309   0.8958
  theta_1     0.274108     0.197397      1.389    0.1650
  theta_2    -0.587919     0.169467     -3.469    0.0005  ***

Mean dependent var   3.883929   S.D. dependent var   3.040381
Mean of innovations -0.053524   S.D. of innovations  1.831760
Log-likelihood      -114.4824   Akaike criterion     240.9648
Schwarz criterion    253.1169   Hannan-Quinn         245.6761

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1           1.1691     0.0000     1.1691     0.0000
    Root  2         -37.7065     0.0000    37.7065     0.5000
  MA
    Root  1          -1.0917     0.0000     1.0917     0.5000
    Root  2           1.5580     0.0000     1.5580     0.0000
  -----------------------------------------------------------
```