**Business School, Brunel University**
**MSc. EC5501/5509 Modelling Financial Decisions and Markets/Introduction to Quantitative Methods**
**Prof. Menelaos Karanasos** (Room SS269, Tel. 01895265284)

**Lecture Notes 3**

# 1. The Classical Linear Regression Model: The Bivariate Case

Suppose that economic theory suggests that $Y = f(X)$, and in particular that $Y$ depends linearly on $X$ :

$$Y = \alpha + \beta X. \tag{1}$$

(You can think of $Y$ and $X$ as representing consumption and income, respectively.) This theoretical model provides the basis for an empirical study. Note that the **deterministic** (exact) relationship (1) serves as a rule and specifies that a unit change in $X$ will produce a $\beta$ units change in $Y$, on average. In order to incorporate in the above model the inherent randomness in its real world counterpart we introduce a stochastic element, $\varepsilon : Y = f(X, \varepsilon)$. The standard approach to incorporate this random element in the relationship between $Y$ and $X$ is to assume that it is additive. So, our **stochastic** equation becomes:

$$Y = \alpha + \beta X + \varepsilon, \tag{2}$$

where $\varepsilon$ is a random **disturbance** (error term). The disturbance (so named because it "disturbs" an otherwise stable relationship) is justified in three main ways:
(i) omission of the influence of innumerable chance effects;
(ii) measurement error (in the dependent variable);
(iii) human indeterminacy (human behaviour can be such that actions taken under identical circumstances will differ in a random way, and so the disturbance term can be thought of as representing this inherent randomness in human behaviour).

The above stochastic equation (2) is the empirical counterpart of the theoretical model (1). We assume that each observation in our sample $(Y_t, X_t)$, $t = 1, .., T$, is generated by an underlying process described by

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \tag{3}$$

where, $\alpha$ and $\beta$ are called *parameters of interest*, $Y$ is the *dependent* (endogenous,or explained, or regressand) variable, $X$ is the *independent* (exogenous, or explanatory, or regressor) variable, and $t$ indexes the $T$ sample observations.

For any set of values of the parameters, $\widehat{\alpha}$ and $\widehat{\beta}$, values of the dependent variable $\left(\widehat{Y}_t\right)$ can be calculated using the values of the independent variable in the data set:

$$\widehat{Y}_t = \widehat{\alpha} + \widehat{\beta} X_t, \tag{4}$$

where $\widehat{Y}_t$, $t = 1, ...T$, are called the **fitted** or predicted values of $Y_t$. Subtraction of the fitted values from the **actual** values $(Y_t)$ produces the **residuals** $\widehat{\varepsilon}_t$, $t = 1, ...T$ (they can be thought of as estimates of the disturbances):

$$Y_t - \widehat{Y}_t = \widehat{\varepsilon}_t, \text{ and so } Y_t = \widehat{\alpha} + \widehat{\beta} X_t + \widehat{\varepsilon}_t. \tag{5}$$

A different set of values of the parameters would create a different estimated line and thus a different set of residuals. It seems natural to ask that a "good" estimator be one that generates a set of estimates that makes these residuals "small". The most popular definition of "small" is the minimisation of the sum of squared residuals. The estimator generating the set of values of the parameters that minimize the sum of squared residuals is called the ordinary least squares estimator (OLS).

**Assumptions of the Bivariate CLRM:**

| | |
|---|---|
| **[A1]** | Functional form: $Y_t$ depends linearly on $X_t$. |
| **[A2]** | Zero mean of the error term: $E(\varepsilon_t) = 0$, for all $t$. |
| **[A3]** | Homoscedasticity: $Var(\varepsilon_t) = \sigma^2$, for all $t$. |
| **[A4]** | No serial correlation: $Cov(\varepsilon_t, \varepsilon_s) = 0$, if $t \neq s$. |
| **[A5]** | $X_t$ is non stochastic. |
| **[A6]** | the error term is normal. |

- Assumption [A5] can be relaxed by assuming that $Cov(X_t, \varepsilon_t) = 0$ (i.e. allowing $X_t$ to be stochastic and uncorrelated with the disturbance).

- Assumptions [A2]-[A4] can be summarised as

$$\varepsilon_t \sim iid\left(0, \sigma^2\right)$$

  which reads: "the disturbances are independently, identically distributed with zero mean and constant variance", whereas assumptions [A2]-[A4] and [A6] can be summarised as

$$\varepsilon_t \sim IN\left(0, \sigma^2\right).$$

- The deterministic part of equation (3) is the expectation of $Y_t$ given $X_t$ :

$$E(Y_t/X_t) = \alpha + \beta X_t.$$

- The variance of the error term is the conditional variance of $Y_t$ on $X_t$ :

$$V(Y_t/X_t) = \sigma^2.$$

  So speaking, the parameters of interest in the context of the bivariate linear regression model are $\alpha, \beta$, and $\sigma^2$.

### 1.1. Deriving the OLS Estimators

One of the several ways to derive the OLS estimators in the context of the bivariate linear regression is the following:

$$\begin{aligned}
S &= \sum_{t=1}^{T}\widehat{\varepsilon}_t^2 = \sum_{t=1}^{T}\left(Y_t - \widehat{\alpha} - \widehat{\beta}X_t\right)^2, \\
\min_{\widehat{\alpha},\widehat{\beta}} S &: \frac{\partial S}{\partial \widehat{\alpha}} = 0 \Rightarrow -2\sum_{t=1}^{T}\left(Y_t - \widehat{\alpha} - \widehat{\beta}X_t\right) = 0 \Rightarrow \\
T\widehat{\alpha} &= \sum_{t=1}^{T} Y_t - \widehat{\beta}\sum_{t=1}^{T} X_t \Rightarrow \\
\widehat{\alpha} &= \overline{Y} - \widehat{\beta}\overline{X}. \qquad (6)
\end{aligned}$$

Substitute the above into (5) to get

$$Y_t = \overline{Y} - \widehat{\beta}\overline{X} + \widehat{\beta}X_t + \widehat{\varepsilon}_t \Rightarrow y_t = \widehat{\beta}x_t + \widehat{\varepsilon}_t, \qquad (5')$$

where $y_t$ and $x_t$ denote the variables in deviation from their means. So we can write:

$$\begin{aligned}
S &= \sum_{t=1}^{T}\widehat{\varepsilon}_t^2 = \sum_{t=1}^{T}\left(y_t - \widehat{\beta}x_t\right)^2, \\
\min_{\widehat{\beta}} S &: \frac{\partial S}{\partial \widehat{\beta}} = 0 \Rightarrow -2\sum_{t=1}^{T} x_t\left(y_t - \widehat{\beta}x_t\right) = 0 \Rightarrow \\
& \sum_{t=1}^{T} x_t y_t - \widehat{\beta}\sum_{t=1}^{T} x_t^2 = 0 \Rightarrow \\
\widehat{\beta} &= \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}, \text{ or } \widehat{\beta} = \frac{\sum_{t=1}^{T}\left(X_t - \overline{X}\right)\left(Y_t - \overline{Y}\right)}{\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2}. \qquad (7)
\end{aligned}$$

Observe that the above estimator is the sample covariance between $X$ and $Y$ divided by the sample variance of the independent variable $X$.

Finally, we compute our estimate of $\sigma^2$ by

$$s^2 = \frac{\sum_{t=1}^{T} \widehat{\varepsilon}_t^2}{T-2},$$

which is called the variance of the regression. It can be shown that $s^2$ is an unbiased estimator of $\sigma^2$.

### 1.2. Gauss-Markov Theorem

Under assumptions [A1]-[A5], the least squares estimators $\left(\widehat{\alpha} \text{ and } \widehat{\beta}\right)$ are BLUE, i.e. best, linear, unbiased estimators.

### 1.3. Expectation and Variance of the OLS Estimators

$$\begin{aligned}
E\left(\widehat{\beta}\right) &= \beta,\ Var\left(\widehat{\beta}\right) = \frac{\sigma^2}{\sum_{t=1}^{T} x_t^2}, \\
E\left(\widehat{\alpha}\right) &= \alpha,\ Var\left(\widehat{\alpha}\right) = \frac{\sigma^2 \sum_{t=1}^{T} X_t^2}{T \sum_{t=1}^{T} x_t^2}, \\
Cov\left(\widehat{\alpha}, \widehat{\beta}\right) &= -\frac{\overline{X}\sigma^2}{\sum_{t=1}^{T} x_t^2}, \\
E\left(s^2\right) &= \sigma^2,\ Var\left(s^2\right) = \frac{2\sigma^4}{T-2}.
\end{aligned}$$

Using $s^2$ instead of $\sigma^2$ in the above formulae, we obtain the sample variances of $\widehat{\alpha}$ and $\widehat{\beta}$.
(<u>Note:</u> in what follows, we will denote the sample standard errors of $\widehat{\alpha}$ and $\widehat{\beta}$ by $s_{\widehat{\alpha}}$ and $s_{\widehat{\beta}}$, respectively.

### 1.4. Goodness of Fit ($R^2$)

The total variation of $Y$ can be decomposed as follows:

$$\underbrace{\sum_{t=1}^{T}\left(Y_t-\overline{Y}\right)^2}_{\text{Total Variation of } Y} = \underbrace{\sum_{t=1}^{T}\left(\widehat{Y}_t-\overline{Y}\right)^2}_{\text{Explained Variation of } Y} + \underbrace{\sum_{t=1}^{T}\left(Y_t-\widehat{Y}_t\right)^2}_{\text{Residual Variation of } Y}$$

We define the goodness of fit (or the coefficient of determination) as

$$R^2 = \frac{\sum_{t=1}^{T}\widehat{y}_t^2}{\sum_{t=1}^{T}y_t^2} = 1 - \frac{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2}{\sum_{t=1}^{T}y_t^2}.$$

It can be shown that $R^2 = r_{xy}^2$, where $r_{xy}$ is the sample correlation between $X$ and $Y$.

## 1.5. Hypothesis Testing

A major group of statistical inference procedures is hypotheses tests. Hypothesis testing is a way of confronting prior beliefs with sample information. The classical testing procedures involve:

- a "null" or maintained hypothesis, $H_0$, which states that the data in the sample have been generated by a particular population;
- an alternative hypothesis, $H_1$;
- a test-statistic, with a known distribution under the null and the alternative, which will enable us to decide whether $H_0$ should be rejected or not. The same test procedure can lead to different conclusions in different samples. So speaking, the test procedure can lead to two different types of error: *type I error*, i.e. rejection of $H_0$ when it is true, and *type II error,* i.e. failure to reject the null when it is false. The probability of a type I error is the *size of the test*, usually denoted $\alpha$; $(1 - \alpha)$ is usually called the *significance level.* The probability of a type II error is denoted $\beta$. The *power of the test* is the probability that it will reject the null when it is false, i.e. $power = 1 - \beta$. The above can be summarised in the following table:

| | Accept $H_0$ | Reject $H_0$ |
|---|---|---|
| $H_0$ is true | ok | **type I error** |
| $H_0$ is false | **type II error** | ok |

  The size of the test is under the control of the analyst, so a fourth ingredient of a testing procedure is

- a specified size of the test , usually $\alpha = 0.05$.

### 1.6. Testing A single Linear Restriction Which Involves One Parameter

As an example consider that, in the context of the bivariate CLRM (eq.(3)) and under the assumption of normality, we wish to test

$$\begin{aligned} H_0 &: \beta = \beta^* \text{ , where } \beta^* \text{ is some constant} \\ H_1 &: \beta \neq \beta^* \end{aligned}$$

We can follow two alternative testing procedures:
**(I)** The sample evidence on $\beta$ is given by $\widehat{\beta}$. The obvious way to measure the distance between the hypothesized value ( $\beta^*$) and the sample evidence $\left(\widehat{\beta}\right)$ is the difference $\left(\widehat{\beta} - \beta^*\right)$. We need to divide the latter by the standard error $\left(s_{\widehat{\beta}}\right)$ of $\widehat{\beta}$, so that our distance measure does not depend on the units of measurement of the data. The resulting test-statistic is called a $t - test$, since it can be shown that, under $H_0$, it follows a t-distribution with $T - 2$ degrees of freedom:

$$t - test = \frac{\widehat{\beta} - \beta^*}{s_{\widehat{\beta}}} \sim t\,(T-2)\,.$$

*Decision rule:* if (negative critical value)< $t - test$ <(positive critical value), then we cannot reject $H_0$. (Note that as the degrees of freedom increase the distribution of the $t - test$ approximates the standard normal.)

The above testing procedure can also be followed to test that the intercept is equal to a particular value.

**(II)** Alternatively, we can impose the restriction on eq. (3), estimate the resulting model, and then use an $F - test$.
*Decision rule:* if $F - statistic$ <critical value, then accept $H_0$.

For example, suppose that we need to test

$$\begin{aligned} H_0 &: \beta = 0 \\ H_1 &: \beta \neq 0. \end{aligned}$$

In this case the $F - test$ is usually called the $F - statistic$ and we have:

$$\begin{aligned}
\text{Unrestricted model } (UR) &: \quad Y_t = \alpha + \beta X_t + \varepsilon_t, \\
\text{Restricted model } (R) &: \quad Y_t = \alpha + u_t, \\
F - statistic &= \frac{\left(\sum_{t=1}^{T} \widehat{u}_t^2 - \sum_{t=1}^{T} \widehat{\varepsilon}_t^2\right)/1}{\left(\sum_{t=1}^{T} \widehat{\varepsilon}_t^2\right)/\left(T-2\right)} \sim F\left(1, T-2\right), \\
F - statistic &= \frac{\left(RRSS - URSS\right)/1}{URSS/\left(T-2\right)} = \left(t - test\right)^2,
\end{aligned}$$

where $URSS$ denotes the unrestricted residual sum of squares, $\sum_{t=1}^{T} \widehat{\varepsilon}_t^2$, and $RRSS$ denotes the restricted residual sum of squares, $\sum_{t=1}^{T} \widehat{u}_t^2$. It can be shown that:

$$F - statistic = \frac{R_{UR}^2\left(T-2\right)}{\left(1 - R_{UR}^2\right)}.$$

In this sense, the $F - statistic$ tests the statistical significance of the coefficient of determination of the bivariate linear regression model (3).

### 1.7. Models with Logs

In econometrics we often estimate models in log-linear form. It is convenient to work with log-linear models for two reasons:

**(i)** The slope coefficient in a log-linear model

$$\log Y_t = \alpha + \beta \log X_t + \varepsilon_t$$

is the elasticity of $Y_t$ with respect to $X_t$, i.e.

$$\varepsilon_{yx} = \left(\frac{dY_t}{dX_t}\right)\left(\frac{X_t}{Y_t}\right) = \frac{d\log Y_t}{d\log X_t} = \beta.$$

**(ii)** The difference of the logarithms of a (time) series is approximately the growth rate ($g$) of the series:

$$\begin{aligned}
Y_t &= \left(1+g\right) Y_{t-1} \Rightarrow \\
\log Y_t &= \log\left(1+g\right) + \log Y_{t-1},
\end{aligned}$$

So

$$\log Y_t - \log Y_{t-1} = \log(1+g).$$

But for small $g$, we have that $\log(1+g)$ is approximately equal to $g$. i.e.

$$\begin{aligned} \log Y_t - \log Y_{t-1} &\cong g, \text{ or} \\ \log Y_t - \log Y_{t-1} &\cong \frac{Y_t - Y_{t-1}}{Y_{t-1}}. \end{aligned}$$

(Note: we use natural logs, i.e. logs to the base $e$.)