1

Lecture 14. The bivariate regression model

14.1. Introduction

So far, we have considered the specification and identification of econometric models. Now we consider ways of estimating the parameters of such models, and ways of testing hypotheses about them. ***Which method of estimation is appropriate, and the properties of the estimators depend on three crucial features of the model:***

(1). whether we are attempting to estimate reduced form or structural equations;

(2). whether the predetermined variables in a given equation are all exogenous, or whether they include lagged endogenous variables;

(3). what assumptions we are making about the error terms in the model eg. can we assume that successive values of a given error term over time are not correlated with each other?

In this lecture, we consider estimation of the parameters of a single reduced form equation, and how to deal with the various possibilities encapsulated in (2) and (3) above. (We will discuss how to estimate the parameters of structural equations in Lecture 18). Before embarking on this material, we need to formalise what we mean by 'endogenous', 'exogenous' and 'predetermined' variables in the context of econometric models. We have been relying on the rather simplistic notion that the endogenous variables are those that are being 'explained' by the equations of the model, while the exogenous variables are those that are 'given' to us a priori, and not explained by the model. In Lecture 13, we said that 'predetermined' variables are either exogenous or lagged endogenous. We now distinguish between these types of variables more formally as follows:
(A). In the context of a static or dynamic structural equation model, an endogenous variable y must be correlated with one or more disturbance terms in the model ie. $E[y_t u_s] \neq 0$ ***for some values of t and s. (B). In the context of a static or dynamic structural equation model, an exogenous variable x is not correlated with any disturbance terms in the model ie.*** $E[x_t u_s] = 0$ ***for all values of t and s. (C). In the context of a dynamic structural equation model, a variable z is predetermined at time t if it is independent of all current and future disturbances in the model ie.*** $E[z_t u_s] = 0$ ***for all*** $s \geq t$.

Note that in the context of dynamic structural equation models, difficulties can arise with lagged endogenous variables if disturbance terms are autocorrelated. ***If*** u_t ***is independent of its own past values*** $u_{t-1}, u_{t-2}, \ldots$, ***then although*** y_{t-1} ***depends on*** u_{t-1}, ***it may be independent of*** u_t ***and can then be regarded as predetermined. However, if*** u_t ***is correlated with*** u_{t-1}, ***then*** y_{t-1} ***and*** u_t ***must be correlated with each other via their common link with*** u_{t-1}, ***and*** y_{t-1} ***cannot therefore be predetermined.*** We will discuss the problems that this presents for estimation in Lecture 16.

You will also need to revise the rules for manipulating expected values given in Lecture 10 Part 1 (page 6). In particular, we will use the rule which says that if X is a random variable, and Y = a + bX where a and b are constants, then E[Y] = a + b[X]. We will also need the one which says that '***the expected value of a sum of random variables is equal to the sum of their expected values***'.

Finally, you need to be completely 'comfortable' (ie. be able to prove) results like the following:

$$\sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y}) = \sum_{t=1}^{n}(X_t - \overline{X})Y_t = \sum_{t=1}^{n} Y_t X_t - n\overline{Y}\overline{X}$$

$$\sum_{t=1}^{n}(X_t - \overline{X})^2 = \sum_{t=1}^{n}(X_t - \overline{X})X_t = \sum_{t=1}^{n} X_t^2 - n\overline{X}^2$$

Results like these come up again and again when discussing the bivariate regression model.

14.2. Estimation of a reduced form bivariate equation

Suppose we have n pairs of observations on two variables X and Y:

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$$

Treating Y as the dependent variable, we could plot these points as in diagram (A) below:

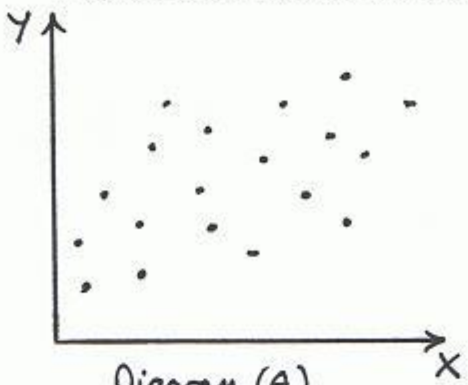


Diagram (A)

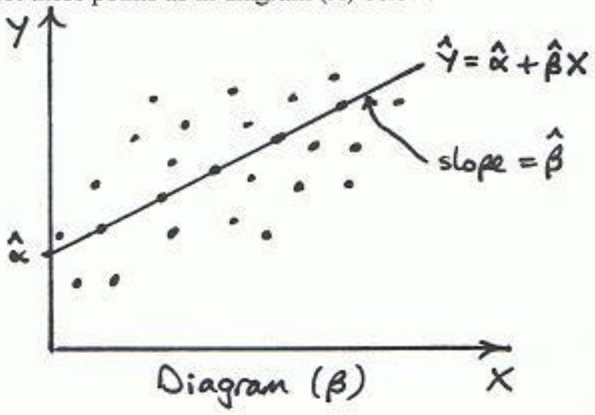


Diagram (B)

In the bivariate regression model, we assume that each Y-value is related to the corresponding X-value according to the following linear equation:

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, \ldots, n$$

where Y_t denotes the tth observation on the dependent variable, X_t the tth observation on the explanatory variable (also called the *regressor*), u_t the tth value of the disturbance term, and α, β are parameters whose values are unknown and are to be estimated. ***We suppose that X_t is a predetermined variable (ie. $E[X_t u_s] = 0$ for $s \geq t$), so that we are dealing with a reduced form equation.***

One way of estimating the parameters α and β is to 'fit' a line to the plotted points by 'visual inspection'. This is illustrated in diagram (B) above. The vertical intercept of the line provides an estimate $\hat{\alpha}$ of the parameter α, and the slope of the line provides an estimate $\hat{\beta}$ of the parameter β. ***The line $\hat{Y} \equiv f(X) = \hat{\alpha} + \hat{\beta}X$ is called the regression of Y on X.*** $\hat{Y}$ is said to be the 'fitted value' of Y corresponding to the value of X. Given an observation X_t of the explanatory variable, we can calculate a fitted value $\hat{Y}_t$ of the dependent variable as follows:

$$\hat{Y}_t \equiv f(X_t) = \hat{\alpha} + \hat{\beta}X_t$$

Thus, given the n observations $X_1, X_2, \ldots, X_n$ on the explanatory variable, we can compute n fitted values $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ of the dependent variable, and compare them to the actually observed values $Y_1, Y_2, \ldots, Y_n$. ***The discrepancies between the observed and fitted values of the dependent variable are called the residuals, denoted by e_t.*** Thus:

$$e_t \equiv Y_t - \hat{Y}_t \qquad t = 1, \ldots, n$$

These can be thought of as 'estimates' of the unobservable disturbances u_t, $t = 1, \ldots, n$.

The 'visual inspection' method of estimating α and β is very subjective; it is unlikely that two different people would construct exactly the same line, so different estimates of α and β could easily arise. A more objective approach, which is also very intuitively appealing, is the statistical method known as Ordinary Least Squares (or OLS), to which we now turn.

Given n pairs of observations on two variables X and Y, ie. $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, the OLS method of estimation involves choosing those estimates $\hat{\alpha}$ and $\hat{\beta}$ which minimise the sum of squared residuals (SSR), defined as

$$SSR \equiv \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2$$

You should know by now that to minimise SSR with respect to $\hat{\alpha}$ and $\hat{\beta}$, we must set the first-order partial derivatives with respect to $\hat{\alpha}$ and $\hat{\beta}$ equal to zero:

$$\frac{\partial SSR}{\partial \hat{\alpha}} = \sum_{t=1}^{n} 2(Y_t - \hat{\alpha} - \hat{\beta} X_t)(-1) = 0$$

$$\frac{\partial SSR}{\partial \hat{\beta}} = \sum_{t=1}^{n} 2(Y_t - \hat{\alpha} - \hat{\beta} X_t)(-X_t) = 0$$

We can rearrange these equations to get

$$\sum_{t=1}^{n} Y_t = n\hat{\alpha} + \hat{\beta} \sum_{t=1}^{n} X_t$$

$$\sum_{t=1}^{n} Y_t X_t = \hat{\alpha} \sum_{t=1}^{n} X_t + \hat{\beta} \sum_{t=1}^{n} X_t^2$$

These are called the OLS normal equations. The normal equations can be solved simultaneously for $\hat{\alpha}$ and $\hat{\beta}$ as follows. Dividing the first normal equation by n gives us

$$\frac{1}{n} \sum_{t=1}^{n} Y_t = \hat{\alpha} + \hat{\beta} \frac{1}{n} \sum_{t=1}^{n} X_t \quad \text{or} \quad \boxed{\overline{Y} = \hat{\alpha} + \hat{\beta} \overline{X}}$$

where $\overline{Y}$ and $\overline{X}$ denote the respective sample means. From the equation $\overline{Y} = \hat{\alpha} + \hat{\beta} \overline{X}$, we get that $\hat{\alpha} = \overline{Y} - \hat{\beta} \overline{X}$. Substituting this into the second normal equation we get

$$\begin{aligned}
\sum_{t=1}^{n} Y_t X_t &= (\overline{Y} - \hat{\beta} \overline{X}) \sum_{t=1}^{n} X_t + \hat{\beta} \sum_{t=1}^{n} X_t^2 \\
&= n(\overline{Y} - \hat{\beta} \overline{X}) \frac{1}{n} \sum_{t=1}^{n} X_t + \hat{\beta} \sum_{t=1}^{n} X_t^2 \\
&= n(\overline{Y} - \hat{\beta} \overline{X}) \overline{X} + \hat{\beta} \sum_{t=1}^{n} X_t^2 \\
&= n\overline{Y}\overline{X} - n\hat{\beta} \overline{X}^2 + \hat{\beta} \sum_{t=1}^{n} X_t^2 \\
&= n\overline{Y}\overline{X} + \hat{\beta} \left(\sum_{t=1}^{n} X_t^2 - n\overline{X}^2 \right)
\end{aligned}$$

Solving this for $\hat{\beta}$ we get

$$\boxed{\hat{\beta} = \frac{\sum_{t=1}^{n} Y_t X_t - n\overline{Y}\overline{X}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2}}$$

This is the OLS estimator for the parameter β in the bivariate regression model. Having found $\hat{\beta}$, we can then calculate $\hat{\alpha}$ from $\hat{\alpha} = \overline{Y} - \hat{\beta}\,\overline{X}$. Thus, we have found the $\hat{\alpha}$ and $\hat{\beta}$ which minimise the SSR (as an exercise, you should check for yourself that the second order conditions for a minimum are satisfied). Let us look more closely at the terms in the numerator and denominator of the OLS estimator $\hat{\beta}$. Looking at the term in the numerator first, we have

$$\sum_{t=1}^{n} Y_t X_t - n\overline{Y}\overline{X} = \sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y}) = (n-1)s_{XY}$$

where

$$s_{XY} \equiv \frac{1}{(n-1)}\sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y})$$

is the <u>sample covariance</u> between X and Y (see Lecture 10 Part 2, page 11). Looking at the term in the denominator we have $\sum_{t=1}^{n} X_t^2 - n\overline{X}^2 = \sum_{t=1}^{n}(X_t - \overline{X})^2 = (n-1)s_X^2$, where

$$s_X^2 \equiv \frac{1}{(n-1)}\sum_{t=1}^{n}(X_t - \overline{X})^2$$

is the <u>sample variance</u> of X. Thus, the OLS estimator $\hat{\beta}$ in the bivariate regression model can be expressed equivalently in several ways:

$$\hat{\beta} = \frac{\sum_{t=1}^{n} Y_t X_t - n\overline{Y}\overline{X}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2} = \frac{\sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y})}{\sum_{t=1}^{n}(X_t - \overline{X})^2} = \frac{(n-1)s_{XY}}{(n-1)s_X^2} = \frac{s_{XY}}{s_X^2}$$

The last equality tells us that $\hat{\beta}$ can be computed as the sample covariance between X and Y divided by the sample variance of X.

14.3. <u>Some implications of the least squares estimates</u>

Some very famous implications of the OLS estimates can be obtained by re-examining the first-order conditions for a minimum of the SSRs, and the normal equations derived from them. We will use some of these later when studying some other aspects of the bivariate regression model. The results we 'prove' in this section are the following:

(a). The method of Ordinary Least Squares always produces estimates $\hat{\alpha}$ and $\hat{\beta}$ which are such that they make the residuals of the regression sum to zero ie. $\sum_{t=1}^{n} e_t = 0$.

(b). The means of the observed and fitted values of Y are equal ie. $\overline{Y} = \overline{\hat{Y}}$.

(c). The method of Ordinary Least Squares always produces estimates $\hat{\alpha}$ and $\hat{\beta}$ which are such that they make the sample covariance between the residuals and the explanatory variable equal to zero ie. $s_{eX} \equiv \frac{1}{(n-1)}\sum_{t=1}^{n}(X_t - \overline{X})e_t = 0$.

(d). The method of Ordinary Least Squares always produces estimates $\hat{\alpha}$ and $\hat{\beta}$ which are such that they make the sample covariance between the residuals and the fitted values of the dependent variable equal to zero ie. $s_{e\hat{Y}} \equiv \frac{1}{(n-1)}\sum_{t=1}^{n}(\hat{Y}_t - \overline{Y})e_t = 0$.

The 'proofs' are as follows:

(a). The method of Ordinary Least Squares always produces estimates $\hat{\alpha}$ and $\hat{\beta}$ which are such that they make the residuals of the regression sum to zero ie. $\sum_{t=1}^{n} e_t = 0$.

This is obvious from the first-order condition $\dfrac{\partial SSR}{\partial \hat{\alpha}} = \sum_{t=1}^{n} 2(Y_t - \hat{\alpha} - \hat{\beta} X_t)(-1) = 0$,

which can be written equivalently as $\sum_{t=1}^{n} (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0$, or $\sum_{t=1}^{n} e_t = 0$.

(b). The means of the observed and fitted values of Y are equal ie. $\overline{Y} = \overline{\hat{Y}}$.

Notice that if $\sum_{t=1}^{n} e_t = 0$, then it must also be true that $\frac{1}{n}\sum_{t=1}^{n} e_t \equiv \bar{e} = 0$. Since $e_t \equiv Y_t - \hat{Y}_t$ for each t, we must have $\bar{e} \equiv \frac{1}{n}\sum_{t=1}^{n} Y_t - \frac{1}{n}\sum_{t=1}^{n} \hat{Y}_t = 0$, which implies that $\overline{Y} = \overline{\hat{Y}}$.

(c). The method of Ordinary Least Squares always produces estimates $\hat{\alpha}$ and $\hat{\beta}$ which are such that they make the sample covariance between the residuals and the explanatory variable equal to zero ie. $s_{eX} \equiv \frac{1}{(n-1)}\sum_{t=1}^{n} (X_t - \overline{X})e_t = 0$.

Consider the other first-order condition. We have $\dfrac{\partial SSR}{\partial \hat{\beta}} = \sum_{t=1}^{n} 2(Y_t - \hat{\alpha} - \hat{\beta} X_t)(-X_t) = 0$

which can be written equivalently as $\sum_{t=1}^{n} (Y_t - \hat{\alpha} - \hat{\beta} X_t)X_t = 0$, or $\sum_{t=1}^{n} e_t X_t = 0$. Now, $\sum_{t=1}^{n} e_t X_t = \sum_{t=1}^{n} (X_t - \overline{X})e_t = (n-1)s_{eX}$, where

$$s_{eX} \equiv \frac{1}{(n-1)}\sum_{t=1}^{n} (X_t - \overline{X})e_t$$

is the sample covariance between the residuals and the explanatory variable X. Thus, if $\sum_{t=1}^{n} e_t X_t = 0$, then it must also be the case that $s_{eX} = 0$.

(d). The method of Ordinary Least Squares always produces estimates $\hat{\alpha}$ and $\hat{\beta}$ which are such that they make the sample covariance between the residuals and the fitted values of the dependent variable equal to zero ie. $s_{e\hat{Y}} \equiv \frac{1}{(n-1)}\sum_{t=1}^{n} (\hat{Y}_t - \overline{Y})e_t = 0$.

Notice that $\hat{Y}_t - \overline{Y} = (\hat{\alpha} + \hat{\beta} X_t) - (\hat{\alpha} + \hat{\beta}\overline{X}) = \hat{\beta}(X_t - \overline{X})$. It follows that

$$s_{e\hat{Y}} \equiv \frac{1}{(n-1)}\sum_{t=1}^{n} (\hat{Y}_t - \overline{Y})e_t = \frac{1}{(n-1)}\sum_{t=1}^{n} \hat{\beta}(X_t - \overline{X})e_t = \hat{\beta}\, s_{eX} = 0.$$

Intuitively, this last result tells us that OLS estimation 'splits' the dependent variable into two components, namely an estimate $\hat{Y} = \hat{\alpha} + \hat{\beta} X$ of the systematic part of Y, and an estimate e of the non-systematic part of Y, in such a way that these two components are uncorrelated.

14.4. Derivation of another expression for $\hat{\beta}$

We have not yet made any assumptions about the disturbance term u_t, because none are needed to get the OLS estimates. ***However, the properties of the estimates vary according to the assumptions made about u_t.*** The most 'desirable' properties are provided by a set of assumptions known as the Classical Assumptions. These are:

(A1). $E[u_t] = 0$ for all t.

(A2). $V[u_t] = E[u_t^2] = \sigma_u^2$ for all t ie. ***the variance of the disturbance term is constant.*** If the constant-variance assumption is satisfied, the disturbances are said to be homoskedastic. If not, they are said to be heteroskedastic.

(A3). $Cov[u_t, u_s] = E[u_t u_s] = 0$ for all $t \neq s$. With time series data, this is the assumption that the errors are not autocorrelated. It is often found to be unrealistic, and we examine the properties of estimates both with and without this assumption.

(A4). X is non-stochastic ie. has no random part. This implies $E[u_t X_s] = X_s E[u_t] = 0$ for all t, s. Note that endogenous variables (whether current or lagged) are stochastic, because they are partly determined by the random disturbances in the model. Thus, assumption (A4) rules out lagged endogenous variables as regressors. Since we are often interested in including lagged endogenous variables in reduced form regressions, we must examine their impact on OLS estimates.

(A5). The regressor exhibits variation ie. the values of X_t, t = 1, . . ., n, are not all the same. Our model is designed to 'explain' the changes in Y that result from changes in X, but if no changes in X are observed, this effect cannot be evaluated.

We now examine the statistical properties of the OLS estimates in various circumstances, focusing on the regression coefficient $\hat{\beta}$. Recall that this can be computed as

$$\hat{\beta} = \frac{\sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

Using the fact that $\sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^{n}(X_t - \bar{X})Y_t$, we can rewrite this as

$$\boxed{\hat{\beta} = \frac{\sum_{t=1}^{n}(X_t - \bar{X})Y_t}{\sum_{t=1}^{n}(X_t - \bar{X})^2}}$$

We study the statistical properties of $\hat{\beta}$ by substituting $Y_t = \alpha + \beta X_t + u_t$ into the above formula to get

$$\hat{\beta} = \frac{\sum_{t=1}^{n}(X_t - \bar{X})(\alpha + \beta X_t + u_t)}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

Separating out the terms in the numerator, we have

$$\hat{\beta} = \frac{\sum_{t=1}^{n}(X_t - \bar{X})\alpha \; + \; \sum_{t=1}^{n}(X_t - \bar{X})\beta X_t \; + \; \sum_{t=1}^{n}(X_t - \bar{X})u_t}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

We can simplify this expression by noting that $\sum_{t=1}^{n}(X_t - \bar{X})\alpha = \alpha\sum_{t=1}^{n}(X_t - \bar{X}) = 0$. We also have $\sum_{t=1}^{n}(X_t - \bar{X})\beta X_t = \beta\sum_{t=1}^{n}(X_t - \bar{X})X_t = \beta\sum_{t=1}^{n}(X_t - \bar{X})^2$. Thus, we can write

$$\hat{\beta} = \frac{\sum_{t=1}^{n}(X_t - \overline{X})(\alpha + \beta X_t + u_t)}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

$$= \frac{\beta\sum_{t=1}^{n}(X_t - \overline{X})^2 + \sum_{t=1}^{n}(X_t - \overline{X})u_t}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

$$= \beta + \frac{\sum_{t=1}^{n}(X_t - \overline{X})u_t}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

To reduce the amount of notation, we define a new variable:

$$\boxed{W_t = \frac{(X_t - \overline{X})}{\sum_{t=1}^{n}(X_t - \overline{X})^2}} \quad t = 1, \ldots, n$$

Using W_t as defined above, we can re-express $\hat{\beta}$ as

$$\boxed{\hat{\beta} = \beta + \sum_{t=1}^{n} W_t u_t}$$

Notice the following very important fact about W_t: it is a function of <u>all</u> the observed values of the explanatory variable X, because its denominator (ie. $\sum_{t=1}^{n}(X_t - \overline{X})^2$) is a sum involving $X_1, X_2, \ldots, X_n$. To fix this idea in our minds, we can write $W_t = f(X_1, X_2, \ldots, X_n)$. As we shall see shortly, this has important implications for OLS estimation when the regressor is a lagged endogenous variable. For future reference, we also note the following properties of W_t:

$$\sum_{t=1}^{n} W_t = 0$$

and

$$\sum_{t=1}^{n} W_t^2 = \frac{\sum_{t=1}^{n}(X_t - \overline{X})^2}{\left\{\sum_{t=1}^{n}(X_t - \overline{X})^2\right\}^2} = \frac{1}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

We will now use the equation $\hat{\beta} = \beta + \sum_{t=1}^{n} W_t u_t$ to examine the following statistical properties of $\hat{\beta}$: (a). Unbiasedness; (b). The variance of $\hat{\beta}$; (c). The consistency of $\hat{\beta}$.

14.4.1. <u>Unbiasedness</u>

We first consider whether the expected value of $\hat{\beta}$ is equal to β ie. whether $E[\hat{\beta}] = \beta$. In this case, $\hat{\beta}$ is an <u>unbiased</u> estimator of β. Using the rules for manipulating expected values (see Lecture 10 Part 1), we have

$$E[\hat{\beta}] = \beta + E[\sum_{t=1}^{n} W_t u_t] = \beta + \sum_{t=1}^{n} E[W_t u_t]$$

Therefore $E[\hat{\beta}] = \beta$ ***(ie. $\hat{\beta}$ is unbiased) if and only if $E[W_t u_t] = 0$ for all t.***

Now, $E[W_t u_t]$ will be zero for all t if X is <u>exogenous</u>, because if X is exogenous, then W must be exogenous too, and by the definition of exogeneity given in the introduction we must have

$$E[W_t u_s] = 0 \text{ for all values of t and s}$$

So if X is <u>exogenous</u>, then $\hat{\beta}$ is an unbiased estimator of β. A special case of this is when X is <u>non-stochastic</u> (assumption (A4) above), because if X is non-stochastic, then W must be non-stochastic too, and we can write

$$E[W_t u_t] = W_t E[u_t] = 0 \text{ for all t}$$

Also notice that it makes no difference whether or not u_t is autocorrelated when X is exogenous, because X is independent of <u>all</u> disturbances in the model, both current and past disturbances.

In contrast, if X is a <u>lagged endogenous</u> variable, say $X_t = Y_{t-1}$, then X_t may be uncorrelated with current and future disturbances (see the definition of predetermined variables in the introduction), but it will <u>not</u> be uncorrelated with u_{t-1}. Remember that, irrespective of the time period t, W_t is a function of <u>all</u> the X's (ie. $W_t = f(X_1, X_2, \ldots, X_n)$) because of the term in its denominator. Therefore if X is a lagged endogenous variable, it must be the case that

$$E[W_t u_t] \neq 0$$

for some t. For example, if $X_t = Y_{t-1}$, then $E[X_t u_{t-1}] \neq 0$, and since W_{t-1} is a function of X_t, we must have

$$E[W_{t-1} u_{t-1}] \neq 0$$

So if X is a <u>lagged endogenous</u> variable, then $\hat{\beta}$ is a <u>biased</u> estimator of β, because $E[\hat{\beta}] \neq \beta$.

Notice that, when X is a lagged endogenous variable, $\hat{\beta}$ will also be biased when the disturbances are autocorrelated.

We can summarise the results of this sub-section in the following table, which you should learn:

Is $\hat{\beta}$ an unbiased estimator of β ?

	u_t not autocorrelated	*u_t autocorrelated*
X_t exogenous	*Yes*	*Yes*
X_t lagged endogenous	*No*	*No*

14.4.2. <u>The variance of</u> $\hat{\beta}$

An unbiased estimator which has a large variance around the true parameter value will be less useful than one which has small variance (ie. one that is more <u>efficient</u>), because with a given sample of data, the one with the smaller variance will be more likely to be near the true parameter value (see the solution to Question 2 (c) in Test 3). The variance of $\hat{\beta}$ is defined as

$$V[\hat{\beta}] = E[(\hat{\beta} - E[\hat{\beta}])^2] = E[(\hat{\beta} - \beta)^2]$$

(see the definition of the variance of a random variable in Lecture 10 Part 1). From the equation

$$\hat{\beta} = \beta + \sum_{t=1}^{n} W_t u_t$$

we get that

$$\hat{\beta} - \beta = \sum_{t=1}^{n} W_t u_t$$

so we can write

$$V[\hat{\beta}] = E[(\hat{\beta} - \beta)^2] = E[(\sum_{t=1}^{n} W_t u_t)^2]$$

When we expand the squared term $(\sum_{t=1}^{n} W_t u_t)^2$ above, we end up with squared terms of the form $(W_t u_t)^2$, and cross-product terms of the form $(W_t u_t)(W_s u_s) = (W_t W_s u_t u_s)$ for $t \neq s$. In total we have

$$V[\hat{\beta}] = E[\sum_{t=1}^{n} (W_t u_t)^2 + \sum_{t \neq s}\sum (W_t W_s u_t u_s)]$$

$$= \sum_{t=1}^{n} E[(W_t u_t)^2] + \sum_{t \neq s}\sum E[W_t W_s u_t u_s]$$

Provided that X (and therefore W) is non-stochastic, we can take W out of the expectation terms, and rewrite the above as

$$V[\hat{\beta}] = \sum_{t=1}^{n} W_t^2 E[u_t^2] + \sum_{t \neq s}\sum W_t W_s E[u_t u_s]$$

If, in addition, assumption (A3) holds (ie. the errors are not autocorrelated), then $E[u_t u_s] = 0$ for $t \neq s$, and if assumption (A2) holds (ie. the errors are homoskedastic), then $E[u_t^2] = \sigma_u^2$ for all t. We then have

$$V[\hat{\beta}] = \sum_{t=1}^{n} W_t^2 \sigma_u^2 + 0 = \sum_{t=1}^{n} W_t^2 \sigma_u^2$$

Finally, remembering that $\sum_{t=1}^{n} W_t^2 = \frac{1}{\sum_{t=1}^{n} (X_t - \bar{X})^2}$, we can write

$$\boxed{V[\hat{\beta}] = \sum_{t=1}^{n} W_t^2 \sigma_u^2 = \frac{\sigma_u^2}{\sum_{t=1}^{n} (X_t - \bar{X})^2}}$$

It can be shown that this variance is the smallest that can be attained by any unbiased estimator under these assumptions. For this reason, the OLS estimator is often called the best linear unbiased estimator (BLUE).

Now, if either of the assumptions (A2) or (A3) fails to hold, the above derivation breaks down. Also, if X is a lagged endogenous variable, then $\hat{\beta}$ is not unbiased (as we saw in the last sub-section), so $\hat{\beta}$ cannot be the BLUE. We can therefore summarise the results of this sub-section in the following table, which you should learn:

Does $V[\hat{\beta}] = \frac{\sigma_u^2}{\sum_{t=1}^{n} (X_t - \bar{X})^2}$, *and is the OLS estimator* $\hat{\beta}$ *the BLUE ?*

	u_t not autocorrelated	*u_t autocorrelated*
X_t exogenous	*Yes*	*No*
X_t lagged endogenous	*No*	*No*

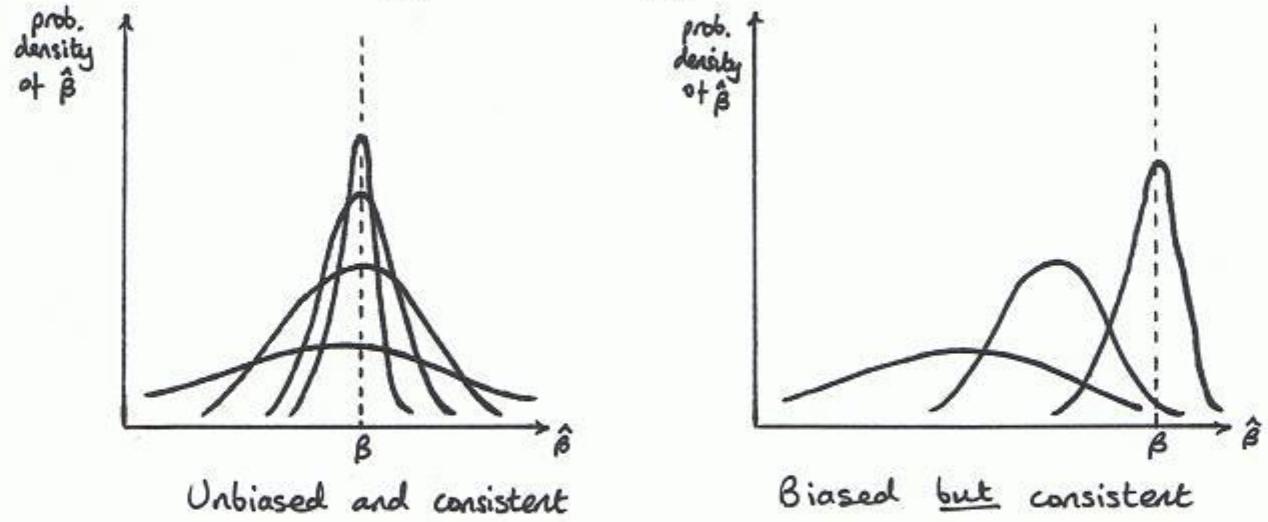
14.4.3. The consistency of $\hat{\beta}$

What can we do if we cannot obtain an unbiased estimator? We now look at an 'asymptotic' property ie. a property that holds in the limiting case as the sample size n tends to infinity. ***If we cannot obtain finite sample results, then the limiting case is the best we can do. This is the situation, for example, when X is a lagged endogenous variable.*** The asymptotic property we consider here is that of consistency (see the solution to Question 2(e) in Test 3). $\hat{\beta}$ ***is a consistent estimator of*** β ***if, as*** $n \to \infty$***, the sampling distribution of*** $\hat{\beta}$ ***concentrates around the true value*** β***, and tends towards a vertical line at*** β ***(called a 'degenerate' distribution).*** In that case, we say that the probability limit of $\hat{\beta}$ is β, and write

$$\text{plim}\,\hat{\beta} = \beta$$

A sufficient condition for consistency is that the bias and variance both tend to zero, that is

$$\lim_{n\to\infty} E[\hat{\beta}] = \beta \quad \text{and} \quad \lim_{n\to\infty} V[\hat{\beta}] = 0$$



This is illustrated in the diagrams above, where the sampling distributions of $\hat{\beta}$ for different sample sizes are shown. These distributions become more and more concentrated around the true value β as the sample size increases, indicating that $\hat{\beta}$ is a consistent estimator.

In the simplest case (X exogenous and non-autocorrelated u's), $\hat{\beta}$ must be consistent, since $E[\hat{\beta}] = \beta$ for all n, and

$$V[\hat{\beta}] = \frac{\sigma_u^2}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \to 0 \quad \text{as } n \to \infty$$

The variance goes to zero because the denominator in the above expression increases without limit as the number of terms, n, goes to infinity (note that all those terms are squares, and therefore positive). It can be shown that $\hat{\beta}$ is also consistent when X is exogenous and the disturbances are autocorrelated.

It can be shown that if X_t is a lagged endogenous variable, then $\hat{\beta}$ is consistent provided that the disturbance terms are not autocorrelated. However, if X_t is a lagged endogenous variable and the disturbance terms are autocorrelated, then $\hat{\beta}$ is inconsistent.

The results on the consistency of $\hat{\beta}$ are summarised in the following table, which you should learn:

Is $\hat{\beta}$ a consistent estimator of β ie. does $plim\,\hat{\beta} = \beta$?

	u_t not autocorrelated	*u_t autocorrelated*
X_t exogenous	*Yes*	*Yes*
X_t lagged endogenous	*Yes*	*No*

14.5. Testing the significance of the coefficients

One of our prime objectives is to test whether X has a significant effect on Y. X has no effect on Y if the true value of the parameter β is zero. ***What we must do is test the null hypothesis that*** β = 0. We write this null hypothesis as

$$H_0: \beta = 0$$

The test developed in this section is based on the sampling distribution of the estimated coefficient $\hat{\beta}$. In order to make statements about this probability distribution, we need an assumption about the probability distribution of the disturbances in the model. We assume that the disturbances are homoskedastic and normally distributed, and write this as

$$u_t \sim N(0, \sigma_u^2) \quad t = 1, \ldots, n$$

Using our equation $\hat{\beta} = \beta + \sum_{t=1}^{n} W_t u_t$, we get that

$$\hat{\beta} - \beta = \sum_{t=1}^{n} W_t u_t$$

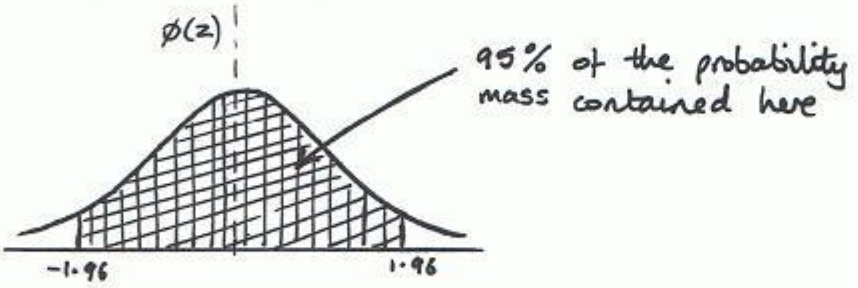
The right hand side of this last expression is a linear combination of normally distributed disturbances, and this means that the term $\hat{\beta} - \beta$ on the left hand side must itself be normally distributed. If X is exogenous and the disturbances are not autocorrelated (so that $\hat{\beta}$ is unbiased and $V[\hat{\beta}] = \frac{\sigma_u^2}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$, as we saw in the previous section), then we have

$$\hat{\beta} - \beta \sim N\left(0,\ \frac{\sigma_u^2}{\sum_{t=1}^{n}(X_t - \bar{X})^2}\right)$$

On dividing by the standard deviation of $\hat{\beta}$ (ie. the square root of $V[\hat{\beta}]$), we obtain a quantity that has the standard normal distribution, which is widely tabulated:

$$\frac{\hat{\beta}-\beta}{\sqrt{\frac{\sigma_u^2}{\sum_{t=1}^{n}(X_t-\overline{X})^2}}} \sim N(0,\,1)$$

If we knew σ_u^2, we could therefore test H_0: $\beta = 0$ using the standard normal distribution. Recall from Lecture 10 Part 2 (page 7) that 95% of the probability mass for a standard normal random variable is contained between the limits -1.96 and 1.96:



If we knew σ_u^2, then imposing the null hypothesis H_0: $\beta = 0$ would yield the following test statistic:

$$\frac{\hat{\beta}}{\sqrt{\frac{\sigma_u^2}{\sum_{t=1}^{n}(X_t-\overline{X})^2}}} \sim N(0,\,1)$$

If the value of this statistic turned out to be greater than 1.96 or less than -1.96, we could say that the probability that it came from a standard normal distribution is less that 0.05, and we could then reject the null hypothesis that $\beta = 0$ ***'at the 5% level of significance'***. Obviously, choosing 'cutoff' points other than ±1.96 would allow us to test the null hypothesis at different significance levels.

Unfortunately, of course, we do not usually know the true value of σ_u^2, and must in practice replace it in the above formula by an unbiased estimate. It can be shown that an unbiased estimate of σ_u^2 is provided by the formula

$$\hat{\sigma}_u^2 = \frac{SSR}{n-2}$$

where $SSR \equiv \sum_{t=1}^{n} e_t^2$ is the sum of squared residuals. Thus, an unbiased estimator of $V[\hat{\beta}] = \frac{\sigma_u^2}{\sum_{t=1}^{n}(X_t-\overline{X})^2}$ is

$$\frac{\hat{\sigma}_u^2}{\sum_{t=1}^{n}(X_t-\overline{X})^2} = \frac{SSR}{(n-2)\sum_{t=1}^{n}(X_t-\overline{X})^2}$$

The square root of this quantity provides an estimate of the standard deviation of the sampling distribution of $\hat{\beta}$ known as the standard error of $\hat{\beta}$ (denoted by $SE(\hat{\beta})$):

$$SE(\hat{\beta}) = \sqrt{\frac{SSR}{(n-2)\sum_{t=1}^{n}(X_t - \overline{X})^2}}$$

Using the fact that $\sum_{t=1}^{n}(X_t - \overline{X})^2 = \sum_{t=1}^{n} X_t^2 - n\overline{X}^2$, we can write this formula in more convenient form for computations:

$$SE(\hat{\beta}) = \sqrt{\frac{SSR}{(n-2)\left\{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2\right\}}}$$

Under the above assumptions, it can be shown that the variable

$$\frac{\hat{\beta} - \beta}{SE(\hat{\beta})}$$

has a t-distribution with n-2 degrees of freedom. This is written as

$$\frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \sim t_{n-2}$$

This quantity can be used for testing hypotheses about β just as before. The only difference is that we now use the t-distribution instead of the standard normal. Imposing the null hypothesis H_0: $\beta = 0$ is true, we get the following <u>test statistic</u>:

$$\frac{\hat{\beta}}{SE(\hat{\beta})} \sim t_{n-2}$$

This quantity is known as the 't-ratio' of $\hat{\beta}$***. To test the null hypothesis*** H_0: $\beta = 0$***, we calculate the t-ratio and compare it with 'cutoff' points from the t-distribution with n-2 degrees of freedom. As a 'rule of thumb', we choose the cutoff points ±2, providing a test at the '5% level of significance' (see below). If the calculated t-ratio is smaller than 2 in absolute value, we do not reject the null hypothesis. If the calculated t-ratio is larger than 2 in absolute value, we reject the null hypothesis. When we reject the null hypothesis, we say that*** $\hat{\beta}$ ***is 'significantly different from zero' (and that X has a significant effect on Y).***

All t-distributions (you get different ones for different 'degrees of freedom') are 'bell-shaped', and look more or less like the standard normal distribution. The difference is that the 'tails' of a particular t-distribution are thicker. However, ***as the sample size n tends to infinity, all t-distributions become more and more like the standard normal distribution.*** Since t-distributions approach the standard normal as $n \to \infty$, the 'cutoff' points for testing 'at the 5% significance level' also approach those of the standard normal distribution ie. ±1.96. If $n > 27$ in the t-tests of $\hat{\beta}$ above, then it turns out that the required 'cutoff' points for the t-distribution are less than 2.06 in absolute value ie. at least 95% of the probability mass for a t-distribution is contained between the limits -2.06 and 2.06 when $n > 27$ in the above tests. ***This explains the simple 'rule of thumb' often used to test*** H_0: $\beta = 0$: ***a coefficient is said to be significant if its <u>absolute value</u> is more than twice its standard error ie. if the absolute value of its t-ratio is greater than 2.*** In reporting empirical work, the usual practice is to present either the standard error, or the t-ratio in parentheses beneath the estimated coefficient. It is important to indicate which is being presented.

We can also use the above procedure to test the hypothesis that β is equal to some other specified value. For example, we might want to test the null hypothesis H_0: β = 1. In this case, we would substitute β = 1 in the above formula to give as our test statistic

$$\frac{\hat{\beta}-1}{SE(\hat{\beta})}$$

Provided that n is not too small, we could again apply the rule of thumb that if this test statistic does not exceed 2 in absolute value, then we cannot reject the null hypothesis H_0: β = 1 at the 5% significance level.

Now we consider the conditions under which the above testing procedure is valid. In devising the test, we have made use of assumptions (A2), (A3) and (A4). So what happens to the test if we have either autocorrelated errors, or X is a lagged dependent variable, or both? ***In the presence of autocorrelation, the expression derived for*** $SE(\hat{\beta})$ ***is no longer correct, and so the test statistic no longer has a t-distribution.*** Thus, the testing procedure breaks down. This is true irrespective of whether X is exogenous or lagged endogenous. ***If X is a lagged endogenous variable, and the disturbances are not autocorrelated, then it can be shown that the test procedure is only valid asymptotically ie. as the sample size tends to infinity.*** With small samples the test can be misleading. The following table summarises the main results. You should learn it:

Is the test procedure outlined above valid ?

	u_t not autocorrelated	***u_t autocorrelated***
X_t exogenous	*Yes*	*No*
X_t lagged endogenous	*Only asymptotically*	*No*

14.6. Goodness of fit

The next question that we briefly examine in this section is how much of the variation in Y is 'explained' by the fitted line $\hat{\alpha}+\hat{\beta}X$. ***We are going to derive a measure of the explanatory power of the equation***. In Sections 14.2 and 14.3 (proposition (d)), we saw that an observed value of the dependent variable Y_t can be 'decomposed' into two uncorrelated parts: a 'fitted' value $\hat{Y}_t \equiv f(X_t) = \hat{\alpha} + \hat{\beta}X_t$, and a 'residual' e_t. Thus, we can write

$$Y_t = \hat{Y}_t + e_t \qquad t = 1, \ldots, n$$

Since the residuals have a zero mean $\bar{e} = 0$ (see proposition (b) in Section 14.3), the means of the observed and fitted values are equal, so the equation

$$(Y_t - \bar{Y}) = (\hat{Y}_t - \bar{Y}) + e_t \qquad t = 1, \ldots, n$$

expresses both the dependent variable and its fitted value in terms of their 'deviations from their means'. Squaring both sides of this equation gives

$$(Y_t - \bar{Y})^2 = (\hat{Y}_t - \bar{Y})^2 + 2(\hat{Y}_t - \bar{Y})e_t + e_t^2 \qquad t = 1, \ldots, n$$

Summing both sides over t gives

$$\sum_{t=1}^{n} (Y_t - \overline{Y})^2 = \sum_{t=1}^{n} (\hat{Y}_t - \overline{Y})^2 + \sum_{t=1}^{n} e_t^2$$

Note that the cross-products $2(\hat{Y}_t - \overline{Y})e_t$ sum to zero since, as shown in proposition (d) in Section 14.3, the covariance between residuals and fitted values is zero. This last equation decomposes the sum of squared 'deviations from the mean' of the dependent variable into the sum of squared deviations from the mean of the fitted values, and the sum of squared deviations from the mean of the residuals (remember, the mean of the residuals is zero). The equation expressed in words is

total variation in Y = explained variation + unexplained variation

If we divided by (n-1) across the equation, we would get the corresponding partition of the variance of the dependent variable.

A natural measure of the 'goodness of fit' or 'explanatory power' of a regression equation is *the proportion of the total variation in Y that is explained by the regression.* This is denoted by R^2 and is called the coefficient of determination:

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = 1 - \frac{\text{unexplained variation}}{\text{total variation}} = 1 - \frac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n} (Y_t - \overline{Y})^2}$$

It can easily be shown that $0 \le R^2 \le 1$, with $R^2 = 1$ only if all the data points lie on the regression line (ie. only if all residuals are zero), and $R^2 = 0$ if X and Y are uncorrelated.

An important point that you should be aware of is that the measure R^2 should not be used if the model does not include a constant term (ie. α). In this case, the meaning of R^2 is not clear, and it is possible for its calculated value to be negative, or to exceed 1. The reason for this is that if there were no constant term included in the regression equation, it would no longer necessarily be true that $\bar{e} = 0$ (recall from Section 14.3 that $\bar{e} = 0$ is a direct implication of the first-order condition $\frac{\partial SSR}{\partial \hat{\alpha}} = \sum_{t=1}^{n} 2(Y_t - \hat{\alpha} - \hat{\beta} X_t)(-1) = 0$, which no longer applies if there is no constant term in the model). But if $\bar{e} = 0$ is not true, then it is no longer true that the means of the observed and fitted values of the dependent variable are both equal to $\overline{Y}$, and the above partition of the sum of squares of mean deviations does not hold.

(End of Lecture 14)

Assignment for Lecture 14. The bivariate regression model

Please make sure that you can do the following problems before next week. You will be asked to solve very similar ones in your exams. Full solutions are attached, but you should not look at these until you have tried your best.

Question 1

16 pairs of observations on Y and X were used to estimate the coefficients in the following equation by the method of Ordinary Least Squares:

$$Y_t = \alpha + \beta X_t + u_t$$

The sum of squared residuals is $\Sigma e_t^2 = 126$. *In addition:*

$$\Sigma X_t Y_t = 492 \qquad \Sigma Y_t = 64 \qquad \Sigma X_t = 96 \qquad \Sigma X_t^2 = 657$$

Showing your workings in full, deduce the estimated values of α *and* β. *Assuming that* $u_t \sim N(0, \sigma_u^2)$, *test the null hypothesis that* $\beta = 1.0$.

Question 2

Consider the following equation:

$$Y_t = \alpha + \beta X_t + u_t$$

Suppose that the parameters α *and* β *are to be estimated by the method of Ordinary Least Squares (OLS). Briefly discuss the implications for the unbiasedness, the variance, and the consistency of the OLS estimator of* β *if*

(a). $X_t = Y_{t-1}$ *and* u_t *is not autocorrelated;*

(b). X_t *is exogenous, but* u_t *follows a first-order autoregressive process of the form* $u_t = \rho u_{t-1} + \varepsilon_t$;

(c). $X_t = Y_{t-1}$, *and* u_t *follows a first-order autoregressive process of the form* $u_t = \rho u_{t-1} + \varepsilon_t$.

Question 3

Briefly discuss the implications for the validity of your test of the null hypothesis that $\beta = 1.0$ *in Question 1 if*

(a). $X_t = Y_{t-1}$ *and* u_t *is not autocorrelated;*

(b). X_t *is exogenous, but* u_t *follows a first-order autoregressive process of the form* $u_t = \rho u_{t-1} + \varepsilon_t$;

(c). $X_t = Y_{t-1}$, *and* u_t *follows a first-order autoregressive process of the form* $u_t = \rho u_{t-1} + \varepsilon_t$.

Assignment for Lecture 14. The bivariate regression model

SOLUTIONS

Question 1

We have $n=16$, $\bar{X} = \frac{96}{16} = 6$, $\bar{Y} = \frac{64}{16} = 4$

So:

$$\hat{\beta} = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} = \frac{492 - (16)(6)(4)}{657 - (16)(36)} = \frac{108}{81} = \frac{4}{3}$$

Then $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 4 - (\frac{4}{3})(6) = -4$

So $\boxed{\hat{\alpha} = -4}$ $\boxed{\hat{\beta} = \frac{4}{3}}$

The test statistic is $\frac{\hat{\beta} - 1}{SE(\hat{\beta})} = \frac{1/3}{SE(\hat{\beta})}$

$$SE(\hat{\beta}) = \sqrt{\frac{SSR}{(n-2)\{\Sigma X^2 - n\bar{X}^2\}}} = \sqrt{\frac{126}{(14)\{657 - (16)(36)\}}}$$

$$= \sqrt{\frac{126}{1134}} = \frac{1}{3}$$

Thus $\frac{\hat{\beta} - 1}{SE(\hat{\beta})} = \frac{1/3}{1/3} = 1 < 2$

Using our 'rule of thumb', we cannot reject the null hypothesis that $\beta = 1$ at the 5% level of significance. (Note: with $n=16$, we need to look at the t-distribution with 14 degrees of freedom. The "correct" cutoff points for a test at the 5% level are ± 2.1448).

Question 2

(a) $\hat{\beta}$ is biased.

$V[\hat{\beta}] \neq \frac{\sigma^2_u}{\Sigma(x-\bar{x})^2}$ and $\hat{\beta}$ is not the BLUE.

$\hat{\beta}$ is consistent.

(b) $\hat{\beta}$ is unbiased.

$V[\hat{\beta}] \neq \frac{\sigma^2_u}{\Sigma(x-\bar{x})^2}$ and $\hat{\beta}$ is not the BLUE.

$\hat{\beta}$ is consistent

(c) $\hat{\beta}$ is biased

$V[\hat{\beta}] \neq \frac{\sigma^2_u}{\Sigma(x-\bar{x})^2}$ and $\hat{\beta}$ is not the BLUE

$\hat{\beta}$ is not consistent

Question 3

(a) Asymptotically valid

(b) Invalid

(c) Invalid

Revision sheet for Lecture 14. The bivariate regression model

For the purposes of examinations, you should learn the following results pertaining to OLS estimation of the model $Y_t = \alpha + \beta X_t + u_t$. It is your responsibility to make sure that you understand where they come from by carefully working through the handout for Lecture 14. If you get stuck, it is your responsibility to do something about it (eg. come and see me).

(a). $\hat{\beta} = \dfrac{\Sigma_{t=1}^{n} Y_t X_t - n\overline{Y}\overline{X}}{\Sigma_{t=1}^{n} X_t^2 - n\overline{X}^2}$ $\qquad \hat{\alpha} = \overline{Y} - \hat{\beta}\,\overline{X}$.

(b). $\dfrac{\hat{\beta}-\beta}{SE(\hat{\beta})} \sim t_{n-2}$ $\qquad SE(\hat{\beta}) = \sqrt{\dfrac{SSR}{(n-2)\left\{\Sigma_{t=1}^{n} X_t^2 - n\overline{X}^2\right\}}}$ $\qquad SSR = \Sigma_{t=1}^{n} e_t^2$

For example, to test the null hypothesis H_0: $\beta = 1$, calculate the ratio $\dfrac{\hat{\beta}-1}{SE(\hat{\beta})}$, and compare the value with the appropriate cutoff points from a t-distribution with n-2 degrees of freedom (approximately ±2 for a test at the 5% level, as long as n is not too small).

(c). The 't-ratio' of $\hat{\beta}$ is defined as $\dfrac{\hat{\beta}}{SE(\hat{\beta})} \sim t_{n-2}$. To test the null hypothesis H_0: $\beta = 0$, we calculate the t-ratio and compare it with appropriate cutoff points from the t-distribution with n-2 degrees of freedom. As a 'rule of thumb', if n is not too small, we choose the cutoff points ±2, providing a test at the '5% level of significance'. If the calculated t-ratio is smaller than 2 in absolute value, we do not reject the null hypothesis. If the calculated t-ratio is larger than 2 in absolute value, we reject the null hypothesis. When we reject the null hypothesis, we say that $\hat{\beta}$ is 'significantly different from zero' (and that X has a significant effect on Y).

(d). *Is* $\hat{\beta}$ *an unbiased estimator of* β *?*

	u_t not autocorrelated	***u_t autocorrelated***
X_t exogenous	*Yes*	*Yes*
X_t lagged endogenous	*No*	*No*

(e). *Does* $V[\hat{\beta}] = \dfrac{\sigma_u^2}{\Sigma_{t=1}^{n}(X_t - \overline{X})^2}$, *and is the OLS estimator* $\hat{\beta}$ *the BLUE ?*

	u_t not autocorrelated	***u_t autocorrelated***
X_t exogenous	*Yes*	*No*
X_t lagged endogenous	*No*	*No*

(f). *Is* $\hat{\beta}$ *a consistent estimator of* β *ie. does plim* $\hat{\beta} = \beta$ *?*

	u_t not autocorrelated	***u_t autocorrelated***
X_t exogenous	*Yes*	*Yes*
X_t lagged endogenous	*Yes*	*No*

(g). *Is the test procedure outlined above valid ?*

	u_t not autocorrelated	***u_t autocorrelated***
X_t exogenous	*Yes*	*No*
X_t lagged endogenous	*Only asymptotically*	*No*