

LECTURE 3

Multiple Regression 1

Covariance and Orthogonality

Let $x = [x_1, x_2, \dots, x_n]'$ be a vector of n random elements. Then its expected value $E(x)$ is simply the vector containing the expected values of the elements:

$$(136) \quad \begin{aligned} E(x) &= [E(x_1), E(x_2), \dots, E(x_n)]' \\ &= [\mu_1, \mu_2, \dots, \mu_n]'. \end{aligned}$$

The variance–covariance matrix or dispersion matrix of x is a matrix $D(x)$ containing the variances and covariances of the elements:

$$(137) \quad D(x) = \begin{bmatrix} V(x_1) & C(x_1, x_2) & \cdots & C(x_1, x_n) \\ C(x_2, x_1) & V(x_2) & \cdots & C(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(x_n, x_1) & C(x_n, x_2) & \cdots & V(x_n) \end{bmatrix}.$$

The variance–covariance matrix is specified in terms of the vector x by writing

$$(138) \quad \begin{aligned} D(x) &= E\left\{ [x - E(x)][x - E(x)]' \right\} \\ &= E\left\{ \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_n - \mu_n \end{bmatrix} [x_1 - \mu_1 \quad x_2 - \mu_2 \quad \cdots \quad x_n - \mu_n] \right\}. \end{aligned}$$

By forming the outer product within the braces, we get the matrix

$$(139) \quad \begin{bmatrix} (x_1 - \mu_1)^2 & (x_1 - \mu_1)(x_2 - \mu_2) & \cdots & (x_1 - \mu_1)(x_n - \mu_n) \\ (x_2 - \mu_2)(x_1 - \mu_1) & (x_2 - \mu_2)^2 & \cdots & (x_2 - \mu_2)(x_n - \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (x_n - \mu_n)(x_1 - \mu_1) & (x_n - \mu_n)(x_2 - \mu_2) & \cdots & (x_n - \mu_n)^2 \end{bmatrix}.$$

3: MULTIPLE REGRESSION 1

On applying the expectation operator to each of the elements, we get the matrix of variances and covariances.

We are interested in comparing the variance–covariance matrix of $x = [x_1, x_2, \dots, x_n]'$ with its empirical counterpart obtained by estimating the variances and covariances from a set of T observations on x . The deviations of the observations about their sample means can be gathered into a matrix $X = [x_{tj} - \bar{x}_j]$ which is written more explicitly as

$$(140) \quad X = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1n} - \bar{x}_n \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2n} - \bar{x}_n \\ \vdots & \vdots & \dots & \vdots \\ x_{T1} - \bar{x}_1 & x_{T2} - \bar{x}_2 & \dots & x_{Tn} - \bar{x}_n \end{bmatrix}.$$

We can easily see that

$$(141) \quad \frac{1}{T} X'X = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1n} \\ S_{21} & S_{22} & \dots & S_{2n} \\ \vdots & \vdots & \dots & \vdots \\ S_{n1} & S_{n2} & \dots & S_{nn} \end{bmatrix},$$

wherein the generic element is the empirical covariance

$$(142) \quad S_{jk} = \frac{1}{T} \sum_{t=1}^T (x_{tj} - \bar{x}_j)(x_{tk} - \bar{x}_k).$$

Now let us reexamine the definition of a covariance. Let x, y be two scalar random variables. Then their covariance is the given by

$$(143) \quad \begin{aligned} C(x, y) &= E \left[\{x - E(x)\} \{y - E(y)\} \right] \\ &= E(xy) - E(x)E(y). \end{aligned}$$

Let $f(x, y) = f(x|y)f(y) = f(y|x)f(x)$ be the joint probability density function of x, y which is expressed as the product of a conditional and a marginal distribution. Here $f(x)$ and $f(y)$ are the marginal distributions of x and y respectively. The expectation of the joint moment of x and y is

$$(144) \quad E(xy) = \int_x \int_y xy f(x, y) dy dx.$$

In the case that x, y are independent, the conditional distributions become marginal distributions and we have $f(x|y) = f(x)$ and $f(y|x) = f(y)$. Therefore

the joint probability density function factorises as $f(x, y) = f(x)f(y)$. It follows that

$$(145) \quad E(xy) = \int_x xf(x)dx \int_y yf(y)dy = E(x)E(y).$$

Thus, in the case of independence, the covariance becomes

$$(146) \quad \begin{aligned} C(x, y) &= E(xy) - E(x)E(y) \\ &= E(x)E(y) - E(x)E(y) = 0. \end{aligned}$$

It is important to understand that, whereas the independence of x, y implies that $C(x, y) = 0$, the reverse is not generally true—the variables can have a zero covariance without being independent. Nevertheless, if x, y have a joint normal distribution, then the condition $C(x, y) = 0$ does imply their independence. Often we are prepared to assume that a set of random variables is normally distributed because this appears to be a reasonable approximation to the truth.

Our next object is to examine the sample analogue of the condition of zero covariance. Let $x = [x_1 - \bar{x}, \dots, x_T - \bar{x}]'$ and $y = [y_1 - \bar{y}, \dots, y_T - \bar{y}]'$ be vectors of the mean-adjusted observations on two random variables taken over T periods. Then the empirical covariance of the observations is just

$$(147) \quad S_{xy} = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y}) = \frac{1}{T} x'y.$$

It follows that $S_{xy} = 0$ if and only if $x'y = 0$ —which is to say if and only if the vectors x and y are orthogonal.

In fact, if the elements of the vectors x and y are continuously distributed over the real line, then we never expect to find an empirical covariance which is precisely zero. For the probability of such an event is infinitesimally small. Nevertheless, if the processes generating the elements of x and y are statistically independent, then we should expect the empirical autocovariance to have a value which tends to zero as the number of observations increases.

The Assumptions of the Classical Linear Model

In order to characterise the properties of the ordinary least-squares estimator of the regression parameters, we make some conventional assumptions regarding the processes which generate the observations.

Let the regression equation be

$$(148) \quad y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon,$$

3: MULTIPLE REGRESSION 1

which is equation (102) again; and imagine, as before, that there are T observations on the variables. Then these can be arrayed in the matrix form of (103) for which the summary notation is

$$(149) \quad y = X\beta + \varepsilon,$$

where $y = [y_1, y_2, \dots, y_T]'$, $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$, $\beta = [\beta_0, \beta_1, \dots, \beta_k]'$ and $X = [x_{tj}]$ with $x_{t0} = 1$ for all t .

The first of the assumptions regarding the disturbances is that they have an expected value of zero. Thus

$$(150) \quad E(\varepsilon) = 0 \quad \text{or, equivalently,} \quad E(\varepsilon_t) = 0, \quad t = 1, \dots, T.$$

Next it is assumed that the disturbances are mutually uncorrelated and that they have a common variance. Thus

$$(151) \quad D(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I \quad \text{or, equivalently,} \quad E(\varepsilon_t\varepsilon_s) = \begin{cases} \sigma^2, & \text{if } t = s; \\ 0, & \text{if } t \neq s. \end{cases}$$

If t is a temporal index, then these assumptions imply that there is no inter-temporal correlation in the sequence of disturbances. In an econometric context, this is often implausible, and we shall relax the assumption at a later stage.

The next set of assumptions concern the matrix X of explanatory variables. A conventional assumption, borrowed from the experimental sciences, is that

$$(152) \quad X \text{ is a nonstochastic matrix with linearly independent columns.}$$

The condition of linear independence is necessary if the separate effects of the k variables are to be distinguishable. If the condition is not fulfilled, then it will not be possible to estimate the parameters in β uniquely, although it may be possible to estimate certain weighted combination of the parameters.

Often, in the design of experiments, an attempt is made to fix the explanatory or experimental variables in such a way that the columns of the matrix X are mutually orthogonal. The device of manipulating only one variable at a time will achieve the effect. The danger of miss-attributing the effects of one variable to another is then minimised.

In an econometric context, it is often more appropriate to regard the elements of X as random variables in their own right, albeit that we are usually reluctant to specify in detail the nature of the processes which generate the variables. Thus we may declare that

$$(153) \quad \begin{array}{l} \text{The elements of } X \text{ are random variables which are} \\ \text{distributed independently of the elements of } \varepsilon. \end{array}$$

The consequence of either of these assumptions (152) or (153) is that

$$(154) \quad E(X'\varepsilon|X) = X'E(\varepsilon) = 0.$$

In fact, for present purposes, it makes little difference which of these assumptions regarding X we adopt; and, since the assumption under (152) is more briefly expressed, we shall adopt it in preference.

The first property to be deduced from the assumptions is that

$$(155) \quad \begin{array}{l} \text{The ordinary least-square regression estimator} \\ \hat{\beta} = (X'X)^{-1}X'y \text{ is unbiased such that } E(\hat{\beta}) = \beta. \end{array}$$

To demonstrate this, we may write

$$(156) \quad \begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= \beta + (X'X)^{-1}X'\varepsilon. \end{aligned}$$

Taking expectations gives

$$(157) \quad \begin{aligned} E(\hat{\beta}) &= \beta + (X'X)^{-1}X'E(\varepsilon) \\ &= \beta. \end{aligned}$$

Notice that, in the light of this result, equation (156) now indicates that

$$(158) \quad \hat{\beta} - E(\hat{\beta}) = (X'X)^{-1}X'\varepsilon.$$

The next deduction is that

$$(159) \quad \begin{array}{l} \text{The variance-covariance matrix of the ordinary least-squares} \\ \text{regression estimator is } D(\hat{\beta}) = \sigma^2(X'X)^{-1}. \end{array}$$

To demonstrate the latter, we may write a sequence of identities:

$$(160) \quad \begin{aligned} D(\hat{\beta}) &= E\left\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\right\} \\ &= E\left\{(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right\} \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\ &= (X'X)^{-1}X'\{\sigma^2I\}X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}. \end{aligned}$$

3: MULTIPLE REGRESSION 1

The second of these equalities follows directly from equation (158).

Statistical Inference and the Assumption of Normality

The dispersion matrix $D(\hat{\beta}) = \sigma^2(X'X)^{-1}$ provides the basis for constructing confidence intervals for the regression parameters in β and for conducting tests of hypotheses relative to these parameters.

For the purposes of statistical inference, it is commonly assumed that the disturbance vector ε has a normal distribution which is denoted by $N(\varepsilon; 0, \sigma^2 I)$. This notation displays the argument of the probability density function together with its expected value $E(\varepsilon) = 0$ and its dispersion $D(\varepsilon) = \sigma^2 I$. The assumption is also conveyed by writing

$$(161) \quad \varepsilon \sim N(0, \sigma^2 I).$$

The assumption implies that the vector y of the dependent variable also has a normal distribution whose mean is the vector $E(y|X) = X\beta$. Thus $y \sim N(X\beta, \sigma^2 I)$. Since the estimator $\hat{\beta}$ is a linear function of y , it follows that it too must have a normal distribution:

$$(162) \quad \hat{\beta} \sim N\{\beta, \sigma^2(X'X)^{-1}\}.$$

In order to specify the distribution of the j th element of $\hat{\beta}$ —which is $\hat{\beta}_j$ —let us denote the j th diagonal element of $(X'X)^{-1}$ by w_{jj} . Then we may assert that

$$(163) \quad \hat{\beta}_j \sim N(\beta_j, \sigma^2 w_{jj}) \quad \text{or, equivalently,} \quad \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 w_{jj}}} \sim N(0, 1).$$

To use this result in making inferences about β_j , we should need to know the value of σ^2 . In its place, we have to make do with an estimate in the form of

$$(164) \quad \hat{\sigma}^2 = \frac{1}{T-k}(y - X\hat{\beta})'(y - X\hat{\beta}) = \frac{e'e}{T-k},$$

which is based on the sum of squares of the residuals.

From our assumption that ε has a normal distribution, it follows that the sum of squares of the residuals, which is $(T-k)\hat{\sigma}^2$, has a chi-square distribution of $T-k$ degrees of freedom. When $\hat{\sigma}^2$ replaces σ^2 in the formulae of (163) we get the result that

$$(165) \quad \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 w_{jj}}} \sim t(T-k),$$

where $t(T - k)$ denotes a t distribution of $T - k$ degrees of freedom.

This result provides the basis for the most common of the inferential procedures in linear regression analysis. In particular, the so-called t -ratios which are displayed in most computer print-outs are simply the quantities $\hat{\beta}_j / \sqrt{\hat{\sigma}^2 w_{jj}}$. We usually declare that the value of the underlying regression parameter is significantly different from zero whenever the value of the t -ratio exceeds 2. [Re. R.J.A. November 23rd].

The assumption that ε has a normal distribution is a particularly convenient theoretical fiction. Even if it is not fulfilled, we can expect $\hat{\beta}$ to have distribution which is approximately a normal distribution. Moreover, the accuracy of this approximation improves as the size of the sample increases. This is the consequence of the remarkable result known as the law of large numbers.

Orthogonality and Omitted-Variables Bias

Let us now investigate the effect that a condition of orthogonality amongst the regressors might have upon the ordinary least-squares estimates of the regression parameters. Let us take the partitioned regression model of equation (109) which was written as

$$(166) \quad y = [X_1, X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

We may assume that the variables in this equation are in deviation form. Let us imagine that the columns of X_1 are orthogonal to the columns of X_2 such that $X_1'X_2 = 0$. This is the same as imagining that the empirical correlation between variables in X_1 and variables in X_2 is zero.

To see the effect upon the ordinary least-squares estimator, we may examine the partitioned form of the formula $\hat{\beta} = (X'X)^{-1}X'y$. Here we have

$$(167) \quad X'X = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} [X_1 \quad X_2] = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix},$$

where the final equality follows from the condition of orthogonality. The inverse of the partitioned form of $X'X$ in the case of $X_1'X_2 = 0$ is

$$(168) \quad (X'X)^{-1} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix}.$$

We also have

$$(169) \quad X'y = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}.$$

3: MULTIPLE REGRESSION 1

On combining these elements, we find that

$$(170) \quad \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1}X_1'y \\ (X_2'X_2)^{-1}X_2'y \end{bmatrix}.$$

In this special case, the coefficients of the regression of y on $X = [X_1, X_2]$ can be obtained from the separate regressions of y on X_1 and y on X_2 .

We should make it clear that this result does not hold true in general. The general formulae for $\hat{\beta}_1$ and $\hat{\beta}_2$ are those which we have given already under (112) and (117):

$$(171) \quad \begin{aligned} \hat{\beta}_1 &= (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2), \\ \hat{\beta}_2 &= \{X_2'(I - P_1)X_2\}^{-1}X_2'(I - P_1)y, \quad P_1 = X_1(X_1'X_1)^{-1}X_1'. \end{aligned}$$

We can easily confirm that these formulae do specialise to those under (170) in the case of $X_1'X_2 = 0$.

The purpose of including X_2 in the regression equation when, in fact, our interest is confined to the parameters of β_1 is to avoid falsely attributing the explanatory power of the variables of X_2 to those of X_1 .

Let us investigate the effects of erroneously excluding X_2 from the regression. In that case, our estimate will be

$$(172) \quad \begin{aligned} \tilde{\beta}_1 &= (X_1'X_1)^{-1}X_1'y \\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon) \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon. \end{aligned}$$

On applying the expectations operator to these equations, we find that

$$(173) \quad E(\tilde{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2,$$

since $E\{(X_1'X_1)^{-1}X_1'\varepsilon\} = (X_1'X_1)^{-1}X_1'E(\varepsilon) = 0$. Thus, in general, we have $E(\tilde{\beta}_1) \neq \beta_1$, which is to say that $\tilde{\beta}_1$ is a biased estimator. The only circumstances in which the estimator will be unbiased are when either $X_1'X_2 = 0$ or $\beta_2 = 0$. In other circumstances, the estimator will suffer from a problem which is commonly described as *omitted-variables bias*.

We need to ask whether it matters that the estimated regression parameters are biased. The answer depends upon the use to which we wish to put the estimated regression equation. The issue is whether the equation is to be used simply for predicting the values of the dependent variable y or whether it is to be used for some kind of structural analysis.

If the regression equation purports to describe a structural or a behavioral relationship within the economy, and if some of the explanatory variables on

the RHS are destined to become the instruments of an economic policy, then it is important to have unbiased estimators of the associated parameters. For these parameters indicate the leverage of the policy instruments. Examples of such instruments are provided by interest rates, tax rates, exchange rates and the like.

On the other hand, if the estimated regression equation is to be viewed solely as a predictive device—that is to say, if it is simply an estimate of the function $E(y|x_1, \dots, x_k)$ which specifies the conditional expectation of y given the values of x_1, \dots, x_n —then, provided that the underlying statistical mechanism which has generated these variables is preserved, the question of the unbiasedness of the regression parameters does not arise.

Multicollinearity

In econometrics, the endeavour to estimate structural parameters is often thwarted by the fact that economic variables are collinear. That is to say, the variables tend to move together, to track one another or to display similar time trends. In the experimental sciences, we can often design an experiment in such a way that the experimental or explanatory variables are uncorrelated or orthogonal. In economics, we rarely have such opportunities.

Let us examine the problem of collinearity within the context of the equation

$$(174) \quad y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t.$$

If we take the data in deviation form, then we can obtain the estimates $\hat{\beta}_1, \hat{\beta}_2$ by solving the following set of equations { Re. R.J.A. October 25th }:

$$(175) \quad \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} S_{1y} \\ S_{2y} \end{bmatrix},$$

wherein

$$(176) \quad S_{ij} = \frac{1}{T} \sum_{t=1}^T (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j).$$

For ease of notation, let us define

$$(177) \quad S_1 = \sqrt{S_{11}} \quad \text{and} \quad S_1^2 = S_{11}.$$

Likewise, we may define S_2 . The empirical correlation coefficient for x_1 and x_2 can then be expressed as

$$(178) \quad r = \frac{S_{12}}{\sqrt{S_{11}S_{22}}}, \quad \text{whence} \quad S_{12} = rS_1S_2.$$

3: MULTIPLE REGRESSION 1

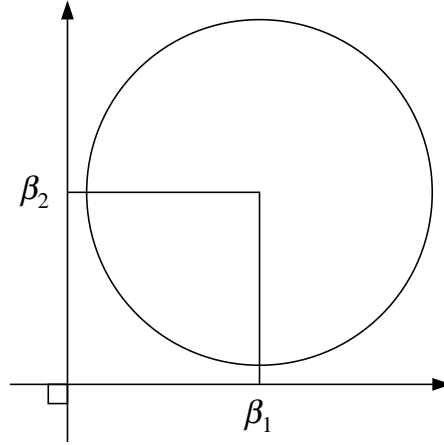


Figure 1.

The correlation coefficient r is a measure of the relatedness of x_1 and x_2 which varies between $+1$ and -1 . The value $r = 1$ indicates a perfect positive linear relationship between the variables whereas the value $r = -1$ indicates a perfect negative relationship.

Now consider the matrix $D(\beta) = \sigma^2(X'X)^{-1}$ of the variances and covariances of the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$. We have $(X'X)^{-1} = \{TS\}^{-1}$, where

$$\begin{aligned}
 (179) \quad S^{-1} &= \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}^{-1} = \begin{bmatrix} S_1^2 & rS_1S_2 \\ rS_2S_1 & S_2^2 \end{bmatrix}^{-1} \\
 &= \frac{1}{S_1^2S_2^2(1-r^2)} \begin{bmatrix} S_2^2 & -rS_1S_2 \\ -rS_2S_1 & S_1^2 \end{bmatrix}
 \end{aligned}$$

is the inverse of the moment matrix of (175).

In the case of $r = 0$, we would find that

$$(180) \quad V(\hat{\beta}_1) = \frac{1}{T} \frac{\sigma^2}{S_1^2} \quad \text{and} \quad V(\hat{\beta}_2) = \frac{1}{T} \frac{\sigma^2}{S_2^2},$$

which is to say that the variance of an estimated parameter is inversely related to corresponding the signal-to-noise ratio.

Figure 1 illustrates the case where $S_1 = S_2$ and $r = 0$. The circle is a contour of the joint probability density function of $\hat{\beta}_1$ and $\hat{\beta}_2$, which is a normal density function on the supposition that the disturbances $\varepsilon_1, \dots, \varepsilon_T$ are normally distributed. We can imagine, for example, that 95% of the probability mass of the joint distribution falls within this contour. The circularity of the contours indicates that the distribution of $\hat{\beta}_2$ is invariant with respect to the realised value of $\hat{\beta}_1$.

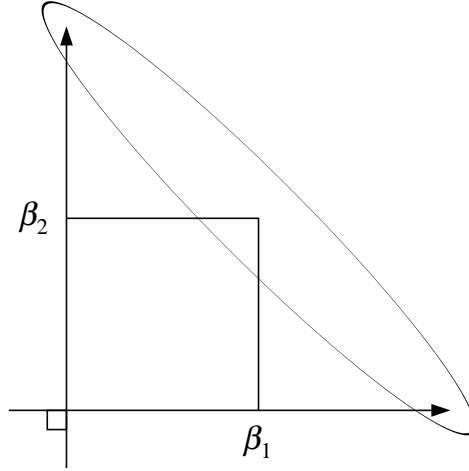


Figure 2.

In general, when $r \neq 0$, we have

$$(181) \quad V(\hat{\beta}_1) = \frac{1}{T} \frac{\sigma^2}{S_1^2(1-r^2)} \quad \text{and} \quad V(\hat{\beta}_2) = \frac{1}{T} \frac{\sigma^2}{S_2^2(1-r^2)},$$

whilst

$$(182) \quad C(\hat{\beta}_1, \hat{\beta}_2) = \frac{1}{T} \frac{-\sigma^2 r}{S_1 S_2 (1-r^2)} \quad \text{and} \quad \text{Corr}(\hat{\beta}_1, \hat{\beta}_2) = \frac{C(\hat{\beta}_1, \hat{\beta}_2)}{\sqrt{V(\hat{\beta}_1)V(\hat{\beta}_2)}} = -r.$$

From this we can see that, as $r \rightarrow 1$, the variances $V(\hat{\beta}_1)$ and $V(\hat{\beta}_2)$ of the two estimates increase without bound whilst their correlation tends to -1 .

Figure 2, depicts the 95% contour of the joint distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$ when $r = 0.75$. The contour is of an elliptical nature which indicates that a knowledge of the realised value of $\hat{\beta}_1$ gives a firm indication of the likely value of $\hat{\beta}_2$. As $r \rightarrow 1$, the ellipse collapses upon the line which is its principal axis.

This result is readily intelligible. Consider the equation

$$(183) \quad y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \varepsilon,$$

and imagine that there is an exact linear relationship between x_1 and x_2 of the form $x_2 = \lambda x_1$ where λ is a constant coefficient. Then we can rewrite the equation as

$$(184) \quad \begin{aligned} y &= \beta_0 + x_1\beta_1 + x_2\beta_2 + \varepsilon \\ &= \beta_0 + x_1(\beta_1 + \lambda\beta_2) + \varepsilon \\ &= \beta_0 + x_1\gamma + \varepsilon, \end{aligned}$$

3: MULTIPLE REGRESSION 1

where $\gamma = (\beta_1 + \lambda\beta_2)$. The set of values of β_1 and β_2 which satisfy this relationship are simply the set of all points on the line defined by the equation $\beta_1 = \gamma - \lambda\beta_2$.