LECTURE 2

Elementary Regression Theory

Regression and Conditional Expectations

Let x and y be a pair of random variables with a well-defined joint probability density function $f(x, y)$. If x is unknown, then the best predictor of y is its unconditional expectation which is defined by

$$(52)\qquad \begin{aligned} E(y) &= \int_y \int_x y f(x, y) dx dy \\ &= \int_y y f(y) dy. \end{aligned}$$

If the value of x is know, then the best predictor is the conditional expectation of y given x which is defined as

$$(53)\qquad \begin{aligned} E(y|x) &= \int_y y \frac{f(x, y)}{f(x)} dy \\ &= \int_y y f(y|x) dy, \end{aligned}$$

where $f(y|x)$ is the conditional probability density function of y given x. The marginal and the conditional expectations are related to each other by the following identity:

$$(54)\qquad E(y) = \int_x E(y|x) f(x) dx.$$

In some cases, it is reasonable to make the assumption that the conditional expectation $E(y|x)$ is a linear function of x:

$$(55)\qquad E(y|x) = \alpha + x\beta.$$

This function is described as a linear regression equation. The error from predicting y by its conditional expectation can be denoted by $\varepsilon = y - E(y|x)$; and therefore we have

$$\text{(56)} \qquad \begin{aligned} y &= E(y|x) + \varepsilon \\ &= \alpha + x\beta + \varepsilon. \end{aligned}$$

Our object is to express the parameters α and β as functions of the moments of the joint probability distribution of x and y. Usually the moments of the distribution can be estimated in a straightforward way from a set of observations on x and y. Using the relationship which exits between the parameters and the theoretical moments, we should be able to find estimates for α and β corresponding to the estimated moments.

We begin by multiplying equation (55) throughout by $f(x)$, and by integrating with respect to x. This gives the equation

$$\text{(57)} \qquad E(y) = \alpha + \beta E(x),$$

whence

$$\text{(58)} \qquad \alpha = E(y) - \beta E(x).$$

Equation (57) shows that the regression line passes through the point $E(x, y) = \{E(x), E(y)\}$ which is the expected value of the joint distribution.

By putting (58) into (55), we find that

$$\text{(59)} \qquad E(y|x) = E(y) + \beta\big\{x - E(x)\big\},$$

which shows how the conditional expectation of y differs from the unconditional expectation in proportion to the error of predicting x by taking its expected value.

Now let us multiply (55) by x and $f(x)$ and then integrate with respect to x to provide

$$\text{(60)} \qquad E(xy) = \alpha E(x) + \beta E(x^2).$$

Multiplying (57) by $E(x)$ gives

$$\text{(61)} \qquad E(x)E(y) = \alpha E(x) + \beta\big\{E(x)\big\}^2,$$

whence, on taking (61) from (60), we get

$$\text{(62)} \qquad E(xy) - E(x)E(y) = \beta\Big[E(x^2) - \big\{E(x)\big\}^2\Big],$$

which implies that

$$\begin{aligned}\beta &= \frac{E(xy) - E(x)E(y)}{E(x^2) - \big\{E(x)\big\}^2} \\ &= \frac{E\Big[\big\{x - E(x)\big\}\big\{y - E(y)\big\}\Big]}{E\Big[\big\{x - E(x)\big\}^2\Big]} \qquad (63) \\ &= \frac{C(x,y)}{V(x)}.\end{aligned}$$

Thus we have expressed α and β in terms of the moments $E(x)$, $E(y)$, $V(x)$ and $C(x, y)$ of the joint distribution of x and y.

It should be recognised that the prediction error $\varepsilon = y - E(y|x) = y - \alpha - x\beta$ is uncorrelated with the variable x. This is shown by writing

$$(64) \qquad E\Big[\big\{y - E(y|x)\big\}x\Big] = E(yx) - \alpha E(x) - \beta E(x^2) = 0,$$

where the final equality comes from (60). This result is readily intelligible; for, if the prediction error were correlated with the value of x, then we should not be using the information of x efficiently in predicting y.

Empirical Regressions

Imagine that we have a sample of T observations on x and y which are $(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T)$. Then we can calculate the following empirical or sample moments:

$$(65) \qquad \bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t,$$

$$(66) \qquad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t,$$

$$(67) \qquad S_x^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})x_t = \frac{1}{T}\sum_{t=1}^{T} x_t^2 - \bar{x}^2,$$

$$(68) \qquad S_{xy} = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y}) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})y_t = \frac{1}{T}\sum_{t=1}^{T} x_t y_t - \bar{x}\bar{y}.$$

It seems reasonable that, in order to estimate α and β, we should replace the moments in the formulae of (58) and (63) by the corresponding sample moments. Thus the estimates of α and β are

$$\begin{aligned}\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x},\\ \hat{\beta} &= \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2}.\end{aligned} \tag{69}$$

The justification of this estimation procedure, which is know as the method of moments, is that, in many of the circumstances under which the sample is liable to be generated, we can expect the sample moments to converge to the true moments of the bivariate distribution, thereby causing the estimates of the parameters to converge likewise to their true values.

Often there is insufficient statistical regularity in the processes generating the variable x to justify our postulating a joint probability density function for x and y. Sometimes the variable is regulated in pursuit of an economic policy in such a way that it cannot be regarded as random in any of the senses accepted by statistical theory. In such cases, we may prefer to derive the estimators of the parameters α and β by methods which make fewer statistical assumptions about x.

When x is a nonstochastic variable, the equation

$$y = \alpha + x\beta + \varepsilon \tag{70}$$

is usually regarded as a functional relationship between x and y which is subject to the effects of a random disturbance term ε. It is commonly assumed that, in all instances of this relationship, the disturbance has a zero expected value and a variance which is finite and constant. Thus

$$E(\varepsilon) = 0 \quad \text{and} \quad V(\varepsilon) = E(\varepsilon^2) = \sigma^2. \tag{71}$$

Also it is assumed that the movements in x are unrelated to those of the disturbance term.

The principle of least squares suggests that we should estimate α and β by finding the values which minimise the quantity

$$\begin{aligned}S &= \sum_{t=1}^{T}(y_t - \hat{y}_t)^2\\ &= \sum_{t=1}^{T}(y_t - \alpha - x_t\beta)^2.\end{aligned} \tag{72}$$

This is the sum of squares of the vertical distances—measured parallel to the y-axis—of the data points from an interpolated regression line.

Differentiating the function S with respect to α and setting the results to zero for a minimum gives

$$(73)\qquad \begin{aligned} -2\sum(y_t-\alpha-\beta x_t)=0, \quad \text{or, equivalently,}\\ \bar{y}-\alpha-\beta\bar{x}=0. \end{aligned}$$

This generates the following estimating equation for α:

$$(74)\qquad \alpha(\beta)=\bar{y}-\beta\bar{x}.$$

Next, by differentiating with respect to β and setting the result to zero, we get

$$(75)\qquad -2\sum x_t(y_t-\alpha-\beta x_t)=0.$$

On substituting for α from (74) and eliminating the factor -2, this becomes

$$(76)\qquad \sum x_t y_t-\sum x_t(\bar{y}-\beta\bar{x})-\beta\sum x_t^2=0,$$

whence we get

$$(77)\qquad \begin{aligned} \hat{\beta}&=\frac{\sum x_t y_t-T\bar{x}\bar{y}}{\sum x_t^2-T\bar{x}^2}\\ &=\frac{\sum(x_t-\bar{x})(y_t-\bar{y})}{\sum(x_t-\bar{x})^2}. \end{aligned}$$

This expression is identical to the one under (69) which we have derived by the method of moments. By putting $\hat{\beta}$ into the estimating equation for α under (74), we derive the same estimate $\hat{\alpha}$ for the intercept parameter as the one to be found under (69).

It is notable that the equation (75) is the empirical analogue of the equation (64) which expresses the condition that the prediction error is uncorrelated with the values of x.

The method of least squares does not automatically provide an estimate of $\sigma^2=E(\varepsilon_t^2)$. To obtain an estimate, we may invoke the method of moments which, in view of the fact that the regression residuals $e_t=y_t-\hat{\alpha}-\hat{\beta}x_t$ represent estimates of the corresponding values of ε_t, suggests an estimator in the form of

$$(78)\qquad \tilde{\sigma}^2=\frac{1}{T}\sum e_t^2.$$

In fact, this is a biased estimator with

$$(79)\qquad E\left(T\tilde{\sigma}^2\right)=\left\{T-2\right\}\sigma^2;$$

so it is common to adopt the unbiased estimator

$$\hat{\sigma}^2 = \frac{\sum e_t^2}{T-2}. \tag{80}$$

The Regression Equation with Two Explanatory Variables

In order to facilitate the treatment of the regression model via matrix algebra, it is useful to recall the algebra of the regression model with two explanatory variables.

Consider the equation

$$y = \alpha + x_1\beta_1 + x_2\beta_2 + \varepsilon, \tag{81}$$

and imagine that there are T observations on y, x_1 and x_2 which are indexed by $t = 1, \ldots, T$. Compared with the former notation, we are using lower-case letters rather than capitals to denote the observations.

According to the principle of least squares, the parameters α, β_1 and β_2 should be estimated by finding the values which minimise the function

$$S = \sum_{t=1}^{T}(y_t - \alpha - x_{t1}\beta_1 - x_{t2}\beta_2)^2. \tag{82}$$

The first-order conditions for the minimisation are obtained by differentiating $S = S(\alpha, \beta_1, \beta_2)$ in respect of its arguments and setting the results to zero. After some trivial simplifications this leads to

$$0 = \sum_t (y_t - \alpha - x_{t1}\beta_1 - x_{t2}\beta_2), \tag{83}$$

$$0 = \sum_t x_{t1}(y_t - \alpha - x_{t1}\beta_1 - x_{t2}\beta_2), \tag{84}$$

$$0 = \sum_t x_{t2}(y_t - \alpha - x_{t1}\beta_1 - x_{t2}\beta_2). \tag{85}$$

On dividing the first of these equations by T are rearranging it, we get the estimating equation for α:

$$\alpha(\beta_1, \beta_2) = \bar{y} - \bar{x}_1\beta_1 - \bar{x}_2\beta_2, \tag{86}$$

where $\bar{x}_1 = T^{-1}\sum_t x_{t1}$ and $\bar{x}_2 = T^{-1}\sum_t x_{t2}$. When this is substituted into the equations (84) and (85) they become

$$0 = \sum_t x_{t1}\Big\{(y_t - \bar{y}) - (x_{t1} - \bar{x}_1)\beta_1 - (x_{t2} - \bar{x}_2)\beta_2\Big\}, \tag{87}$$

$$0 = \sum_t x_{t2}\Big\{(y_t - \bar{y}) - (x_{t1} - \bar{x}_1)\beta_1 - (x_{t2} - \bar{x}_2)\beta_2\Big\}. \tag{88}$$

We can now avail ourselves of a few definitions:

(89) $$S_{11} = \frac{1}{T}\sum_{t=1}^{T}(x_{t1} - \bar{x}_1)^2 = \frac{1}{T}\sum_{t=1}^{T}(x_{t1} - \bar{x}_1)x_{t1},$$

(90) $$S_{22} = \frac{1}{T}\sum_{t=1}^{T}(x_{t2} - \bar{x}_2)^2 = \frac{1}{T}\sum_{t=1}^{T}(x_{t2} - \bar{x}_2)x_{t2},$$

(91) $$S_{12} = \frac{1}{T}\sum_{t=1}^{T}(x_{t1} - \bar{x}_1)(x_{t2} - \bar{x}_2) = \frac{1}{T}\sum_{t=1}^{T}(x_{t1} - \bar{x}_1)x_{t2},$$

(92) $$S_{1y} = \frac{1}{T}\sum_{t=1}^{T}(x_{t1} - \bar{x}_1)(y_t - \bar{y}) = \frac{1}{T}\sum_{t=1}^{T}(x_{t1} - \bar{x}_1)y_t,$$

(93) $$S_{2y} = \frac{1}{T}\sum_{t=1}^{T}(x_{t2} - \bar{x}_2)(y_t - \bar{y}) = \frac{1}{T}\sum_{t=1}^{T}(x_{t2} - \bar{x}_2)y_t.$$

In these terms, the pair of equations under (87) and (88) become

(94) $$S_{11}\beta_1 + S_{12}\beta_2 = S_{1y},$$

(95) $$S_{21}\beta_1 + S_{22}\beta_2 = S_{2y},$$

wherein $S_{21} = S_{12}$. Using simple algebraic manipulations, a solution may be obtained in the form of

(96) $$\hat{\beta}_1 = \frac{S_{1y} - S_{12}\hat{\beta}_2}{S_{11}},$$

(97) $$\hat{\beta}_2 = \frac{S_{11}S_{2y} - S_{12}S_{1y}}{S_{11}S_{22} - S_{12}^2},$$

Alternatively, we may write the equations in a matrix format as

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

Using the formula for the inverse of a matrix of order 2×2, we get

(99) $$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \frac{1}{S_{11}S_{22} - S_{12}^2} \begin{bmatrix} S_{22} & -S_{12} \\ -S_{21} & S_{11} \end{bmatrix} \begin{bmatrix} S_{1y} \\ S_{2y} \end{bmatrix}.$$

On multiplying the vector and the matrix on the RHS we get

(100) $$\hat{\beta}_1 = \frac{S_{22}S_{1y} - S_{12}S_{2y}}{S_{11}S_{22} - S_{12}^2}.$$

together with the expression for $\hat{\beta}_2$ of (97). The estimate of α, which comes from substituting $\hat{\beta}_1$ and $\hat{\beta}_2$ into equation (86), is

$$\hat{\alpha} = \bar{y} - \bar{x}_1\hat{\beta}_1 - \bar{x}_2\hat{\beta}_2. \tag{101}$$

The Multiple Regression Model in Matrices

Consider the regression equation

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon, \tag{102}$$

and imagine that T observations on the variables $y, x_1, \ldots, x_k$ are available which are indexed by $t = 1, \ldots, T$. Then we can write the T realisations of the relationship in the following form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{T1} & \ldots & x_{Tk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}. \tag{103}$$

This can be represented in summary notation by

$$y = X\beta + \varepsilon. \tag{104}$$

Our object is to derive an expression for the ordinary least-squares estimates of the elements of the parameter vector $\beta = [\beta_0, \beta_1, \ldots, \beta_k]'$. The criterion is to minimise a sum of squares of residuals which can be written variously as

$$\begin{aligned} S(\beta) &= \varepsilon'\varepsilon \\ &= (y - X\beta)'(y - X\beta) \\ &= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta \\ &= y'y - 2y'X\beta + \beta'X'X\beta. \end{aligned} \tag{105}$$

Here, to reach the final expression, we have used the identity $\beta'X'y = y'X\beta$ which comes from the fact that the transpose of a scalar—which may be construed as a matrix of order 1×1—is the scalar itself.

To find the first-order conditions, we differentiate the function with respect to the vector β and we set the result to zero. According to the rules of matrix differentiation, which are easily verified, the derivative is

$$\frac{\partial S}{\partial \beta} = -2y'X + 2\beta'X'X. \tag{106}$$

Setting this to zero gives $0 = \beta' X' X - y' X$, which is transposed to provide the so-called normal equations:

$$X'X\beta = X'y. \tag{107}$$

On the assumption that the inverse matrix exists, the equations have a unique solution which is the vector of ordinary least-squares estimates:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{108}$$

The Partitioned Regression Model

Consider taking the regression equation of (104) in the form of

$$y = [\, X_1 \quad X_2 \,] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{109}$$

Here, $[X_1, X_2] = X$ and $[\beta_1', \beta_2']' = \beta$ are obtained by partitioning the matrix X and vector β in a conformable manner. The normal equations of (107) can be partitioned likewise. Writing the equations without the surrounding matrix braces gives

$$X_1'X_1\beta_1 + X_1'X_2\beta_2 = X_1'y, \tag{110}$$

$$X_2'X_1\beta_1 + X_2'X_2\beta_2 = X_2'y. \tag{111}$$

From (110), we get the equation $X_1'X_1\beta_1 = X_1'(y - X_2\beta_2)$ which gives an expression for the leading subvector of $\hat{\beta}$:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2). \tag{112}$$

To obtain an expression for $\hat{\beta}_2$, we must eliminate β_1 from equation (111). For this purpose, we multiply equation (110) by $X_2'X_1(X_1'X_1)^{-1}$ to give

$$X_2'X_1\beta_1 + X_2'X_1(X_1'X_1)^{-1}X_1'X_2\beta_2 = X_2'X_1(X_1'X_1)^{-1}X_1'y. \tag{113}$$

When the latter is taken from equation (111), we get

$$\Big\{X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\Big\}\beta_2 = X_2'y - X_2'X_1(X_1'X_1)^{-1}X_1'y. \tag{114}$$

On defining

$$P_1 = X_1(X_1'X_1)^{-1}X_1', \tag{115}$$

can we rewrite (114) as

$$\left\{X_2'(I-P_1)X_2\right\}\beta_2 = X_2'(I-P_1)y, \tag{116}$$

whence

$$\hat{\beta}_2 = \left\{X_2'(I-P_1)X_2\right\}^{-1} X_2'(I-P_1)y. \tag{117}$$

The Matrix Form for Simple Regression

Now consider again the equations

$$y_t = \alpha + x_t\beta + \varepsilon_t, \quad t = 1, \ldots, T \tag{118}$$

which comprise T observations of the simple regression model. To represent these in a matrix form, we must define the following vectors:

$$\begin{aligned} y &= [y_1, y_2, \ldots, y_T]', \\ x &= [x_1, x_2, \ldots, x_T]', \\ \varepsilon &= [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T]', \\ i &= [1, 1, \ldots, 1]'. \end{aligned} \tag{119}$$

Here the vector $i = [1, 1, \ldots, 1]'$, which consists of T units, is described alternatively as the dummy vector or the summation vector.

In terms of the vector notation, the equation of (118) can be written as

$$y = i\alpha + x\beta + \varepsilon, \tag{120}$$

which can be construed as a case of the partitioned regression equation of (109). By setting $X_1 = i$ and $X_2 = x$ and by taking $\beta_1 = \alpha$, $\beta_2 = \beta$ in equations (112) and (117), we derive the following expressions for the estimates of the parameters α, β:

$$\hat{\alpha} = (i'i)^{-1}i'(y - x\hat{\beta}), \tag{121}$$

$$\begin{aligned} \hat{\beta} &= \left\{x'(I-P_i)x\right\}^{-1} x'(I-P_i)y, \quad \text{with} \\ P_i &= i(i'i)^{-1}i' = \frac{1}{T}ii'. \end{aligned} \tag{122}$$

To understand the effect of the operator P_i in this context, consider the following expressions:

$$
\begin{aligned}
i'y &= \sum_{t=1}^{T} y_t, \\
(i'i)^{-1}i'y &= \frac{1}{T}\sum_{t=1}^{T} y_t = \bar{y}, \\
P_i y &= i(i'i)^{-1}i'y = [\bar{y}, \bar{y}, \ldots, \bar{y}]'.
\end{aligned} \tag{123}
$$

Here $P_i y = [\bar{y}, \bar{y}, \ldots, \bar{y}]'$ is simply a column vector containing T repetitions of the sample mean. From the expressions above, it can be be understood that, if $x = [x_1, x_2, \ldots x_T]'$ is vector of T elements, then

$$
x'(I - P_i)x = \sum_{t=1}^{T} x_t(x_t - \bar{x}) = \sum_{t=1}^{T}(x_t - \bar{x})x_t = \sum_{t=1}^{T}(x_t - \bar{x})^2. \tag{124}
$$

The final equality depends upon the fact that $\sum(x_t - \bar{x})\bar{x} = \bar{x}\sum(x_t - \bar{x}) = 0$.

On using the results under (123) and (124) in the equations (121) and (122), we find that

$$
\hat{\alpha} = \bar{y} - \bar{x}\hat{\beta}, \tag{125}
$$

$$
\hat{\beta} = \frac{\sum_t (x_t - \bar{x})y_t}{\sum_t (x_t - \bar{x})x_t} = \frac{\sum_t (x_t - \bar{x})(y_t - \bar{y})}{\sum_t (x_t - \bar{x})^2}, \tag{126}
$$

which are the formulae to be found under (69).

The Regression Model in Deviation Form

The estimator for β under (126) comprises the deviations of the original observations $x_1, \ldots, x_T$ from their sample mean $\bar{x}$. Also, we are free to replace the observations $y_1, \ldots, y_T$ by their deviations from the corresponding sample mean $\bar{y}$. It follows that the estimate of β is precisely the value which would be obtained by applying the technique of least-squares regression to a meta-equation

$$
y_t - \bar{y} = (x_t - \bar{x})\beta + (\varepsilon_t - \bar{\varepsilon}), \tag{127}
$$

which lacks an intercept term. The estimate for the intercept term can be recovered from the equation (125) once the value for $\hat{\beta}$ is available.

This approach is applicable to equations with any number of explanatory variables. Consider replacing the equation of (103) by the equation

$$(128)\qquad \begin{bmatrix} y_1-\bar{y} \\ y_2-\bar{y} \\ \vdots \\ y_T-\bar{y} \end{bmatrix} = \begin{bmatrix} x_{11}-\bar{x}_1 & \ldots & x_{1k}-\bar{x}_k \\ x_{21}-\bar{x}_1 & \ldots & x_{2k}-\bar{x}_k \\ \vdots & & \vdots \\ x_{T1}-\bar{x}_1 & \ldots & x_{Tk}-\bar{x}_k \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1-\bar{\varepsilon} \\ \varepsilon_2-\bar{\varepsilon} \\ \vdots \\ \varepsilon_T-\bar{\varepsilon} \end{bmatrix}.$$

If we define the matrix $X = [x_{tj} - \bar{x}_j]$ and the vectors $y = [y_t - \bar{y}]$ and $\varepsilon = [\varepsilon_t - \bar{\varepsilon}]$, then we can retain the summary notation $y = X\beta + \varepsilon$ which now denotes equation (128) instead of equation (103).

As an example of this device, let us consider the equation

$$(129)\qquad y_t = \alpha + x_{t1}\beta_1 + x_{t2}\beta_2 + \varepsilon_t, \quad t = 1, \ldots, T,$$

which was displayed, in slightly different notation, in the lecture of November 24th. Compared with the former notation, we are now now setting $\alpha = \beta_0$ and we are using lower-case letters rather than capitals to denote the observations. In the former notation, lower-case letters were used to denote deviations.

The present equation gives rise to the following deviation form:

$$(130)\qquad y_t - \bar{y} = (x_{t1} - \bar{x}_1)\beta_1 + (x_{t2} - \bar{x}_2)\beta_2 + (\varepsilon_t - \bar{\varepsilon}), \quad t = 1, \ldots, T.$$

Let us define the corresponding vectors:

$$(131)\qquad \begin{aligned} y &= [y_1 - \bar{y}, \ldots, y_T - \bar{y}]', \\ x_1 &= [x_{11} - \bar{x}_1, \ldots, x_{T1} - \bar{x}_1]', \\ x_2 &= [x_{12} - \bar{x}_2, \ldots, x_{T2} - \bar{x}_2]', \\ \varepsilon &= [\varepsilon_1 - \bar{\varepsilon}, \ldots, \varepsilon_T - \bar{\varepsilon}]'. \end{aligned}$$

Then the summary notation for the equation (130) is just

$$(132)\qquad y = x_1\beta_1 + x_2\beta_2 + \varepsilon,$$

which is equation (109) with $X_1 = x_1$ and $X_2 = x_2$ and with β_1, β_2 as scalars rather than vectors. It follows that equations (112) and (117) provide the appropriate means of estimating the regression parameters.

With $P_1 = x_1(x_1'x_1)^{-1}x_1'$, we get

$$(133)\qquad \begin{aligned} x_2'(1 - P_1)x_2 &= x_2'x_2 - x_2'x_1(x_1'x_1)^{-1}x_1'x_2 \\ &= T\big\{S_{22} - S_{21}S_{11}^{-1}S_{12}\big\}, \end{aligned}$$

where $S_{21} = S_{12}$, since these are scalars. It follows that

$$\begin{aligned}\hat{\beta}_1 &= (x_1'x_1)^{-1}x_1'(y - x_2\hat{\beta}_2)\\ &= S_{11}^{-1}\{S_{1y} - S_{12}\hat{\beta}_2\},\end{aligned} \tag{134}$$

and that

$$\begin{aligned}\hat{\beta}_2 &= \{x_2'(1 - P_1)x_2\}^{-1}x_2'(1 - P_1)y\\ &= \{S_{22} - S_{21}S_{11}^{-1}S_{12}\}^{-1}\{S_{2y} - S_{21}S_{11}^{-1}S_{1y}\}.\end{aligned} \tag{135}$$

These are the matrix versions of the formulae which have already appeared under (96) and (97).