Derivation of the Ordinary Least Squares Estimator
Multiple Regression Case

In the previous reading assignment the ordinary least squares (OLS) estimator for the simple linear regression case, only one independent variable (only one x), was derived. The procedure relied on combining calculus and algebra to minimize of the sum of squared deviations. The simple linear case although useful in illustrating the OLS procedure is not very realistic. Rarely, are you interested in only one independent variable's potential affect on the dependent variable, y. For example, if interested forecasting sales for your would you find only own price in the econometric forecasting equation(s). Other factors such as competitors' price and the general state of the economy will affect your sales. This more realistic case, more than one independent variable, is the subject of this reading assignment. The OLS estimator is derived for the multiple regression case. Multiple regression simply refers to the inclusion of more than one independent variable.

Problem Set-up

N-Paired Observations

Similar to the simple linear regression problem, you have N-paired observations. In this case for each y observation, there is an associated set of x's. *Key point:* the paired observations are one y associated with a set of x's. As an example, the first three N-paired observations may be

$$
\begin{array}{ccccc}
\underline{y_i} & \underline{x_1} & \underline{x_3} & \cdots & \underline{x_k} \\
45 & 300 & 0 & \cdots & 10 \\
100 & 300 & 15 & \cdots & 5 \\
80 & 280 & 10 & \cdots & 15
\end{array}
$$

(1)

where y_i and x_i represent paired observations and there are k independent variables. Note, the only difference from the simple linear regression case is the addition of independent variables. Observations must still be paired.

Linear Equation

Again, similar to the simple linear regression case, OLS is used to estimate linear equations. The equations must be linear in the parameters. Recall, from the simple linear regression case, linear in parameters refers to linear in the unknown parameters and not linear in the x's. This point will be covered in more detail later in the class. *Key point:* linear in parameters is the assumption made. This assumption is the same as was made in the simple linear regression case. The only difference is now there are more parameters.

Considering only the three above observations and only the three x's, three linear equations are given as

$$45 = \beta_1 + \beta_2 300 + \beta_3 0 + \beta_4 10 + u_1$$

(2)
$$100 = \beta_1 + \beta_2 300 + \beta_3 15 + \beta_4 5 + u_2$$

$$80 = \beta_1 + \beta_2 280 + \beta_3 10 + \beta_4 15 + u_3$$

where the β's are unknown parameters and the u's are the error or residual terms. *Key points are:* 1) there is one equation for each paired-observation and 2) there is an error term associated with each equation. These key points are the same as in the simple linear case.

In general, the multiple regression case can be written as

(3)
$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \cdots + \beta_k x_{tk} + u_t$$

where the β's are k unknown parameters, the u's are the error or residual terms, t refers to the observation number, and x_{ti} refers to the ith independent variable for observation t. Notice the numbering of the x variables begins with two. Implicitly, β_1, the intercept parameter, is multiplied by one. This value of one for x_{t1} is usually omitted when writing the equation. However, this value for x_{t1} becomes very important later in the derivation. The individual equations for k parameters and n observations are

$$y_1 = \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_k x_{1k} + u_1$$

$$y_2 = \beta_1 + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_k x_{2k} + u_2$$

(4)
$$y_3 = \beta_1 + \beta_2 x_{32} + \beta_3 x_{33} + \cdots + \beta_k x_{3k} + u_3 \ .$$

$$\vdots \qquad\qquad \vdots$$

$$y_n = \beta_1 + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_k x_{nk} + u_n$$

The difference between the simple linear and multiple linear case is the complicating issue of additional independent variables, x's. This complicating issue is not trivial. Using the procedure to derive the simple linear case, the derivation of the OLS estimator results in k equations (a FOC for each unknown). Each equation will have n components (a squared error for each observation) before simplification. Writing the equations becomes a time consuming and tedious, basically a nightmare. A short hand notation is necessary. This is where matrix algebra enters the mix. Three matrix algebra operations are necessary, multiplication, transpose, and inverse. All three operations have been covered in the class prerequisites and reviewed earlier. The following discussion also provides sidelights to help in the understanding.

Key point: the derivation of the OLS estimator in the multiple linear regression case is the same as in the simple linear case, except matrix algebra instead of linear algebra is used. Nothing new is added, except addressing the complicating factor of additional independent variables.

Equations in Matrix Form

To write equation (4) in matrix form, four matrices must be defined, one for the dependent variables, one for the independent variables, one for the unknown parameters, and finally one for the error terms. These for matrices are:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad U = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix}, \qquad \text{and}$$

(5)

$$X = \begin{bmatrix} 1 & x_{12} & x_{13} & x_{14} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & x_{24} & \cdots & x_{2k} \\ 1 & x_{32} & x_{33} & x_{34} & \cdots & x_{3k} \\ & & \vdots & & & \\ 1 & x_{n2} & x_{n3} & x_{n4} & \cdots & x_{nk} \end{bmatrix}.$$

Y and U are column vectors of dimension n x 1, β is column vector of dimension k x 1, and X is a matrix of dimension n x k. Using these matrices, equation (4) can be written as

(6)
$$\underset{(n \times 1)}{Y} = \underset{(n \times k)}{X} \underset{(k \times 1)}{\beta} + \underset{(n \times 1)}{U}.$$

It is clear that equation (6) is much simpler to write than writing the equations in equation (4). Further, matrix algebra allows the manipulation of equation (6) much as if it was a single equation. This simplifies the derivation of the multiple linear regression case. Three points one should recognize in equation (6) are:

1) each row corresponds to an individual observation,
2) the column of ones in the X matrix represent the intercept term, and
3) without subscripts the notation denotes a matrix, whereas with subscripts the notation denotes elements of a matrix.

Example of equation (6). Using the observations given in equation (2), the formulation of equation (6) can be shown. To show the formulation, matrix multiplication must be used. Recall, in matrix multiplication the element of the resulting matrix in found by summing the multiplication of the elements in the rows of the first matrix by the corresponding element in the columns of the second matrix.

Using the three observations in equation (2), the appropriate matrices are

$$Y = \begin{bmatrix} 45 \\ 100 \\ 80 \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \qquad U = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \qquad \text{and}$$

$$X = \begin{bmatrix} 1 & 300 & 0 & 10 \\ 1 & 300 & 15 & 5 \\ 1 & 28 & 10 & 15 \end{bmatrix}.$$

Equation (6) is

$$Y + X\beta + U$$

$$\begin{bmatrix} 45 \\ 100 \\ 80 \end{bmatrix} = \begin{bmatrix} 1 & 300 & 0 & 10 \\ 1 & 300 & 15 & 5 \\ 1 & 280 & 10 & 15 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

$$\begin{bmatrix} 45 \\ 100 \\ 80 \end{bmatrix} = \begin{bmatrix} 1\beta_1 + 300\beta_2 + 0\beta_3 + 10\beta_4 \\ 1\beta_1 + 300\beta_2 + 15\beta_3 + 5\beta_4 \\ 1\beta_1 + 280\beta_2 + 10\beta_3 + 15\beta_4 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

Equation (2) is obtained from equation (6) by multiplying out the matrices and using the definition of matrix addition:

$$45 = \beta_1 + \beta_2 300 + \beta_3 0 + \beta_4 10 + u_1$$
$$100 = \beta_1 + \beta_2 300 + \beta_3 15 + \beta_4 5 + u_2 \,.$$
$$80 = \beta_1 + \beta_2 280 + \beta_3 10 + \beta_4 15 + u_3$$

By example, we have shown that equation (4) is much simpler to write than equation (2). Further, equation (4) will be simpler to manipulate.

Next, we need to define the estimated error associated with each observation and put the error into matrix form. As in the simple linear case, for any observation, the estimated error term is defined as the actual y value minus the estimated y value $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{a} - \hat{b}x_i$. To obtain the error term, the unknown values for the parameters is replaced by the estimated values. In the multiple linear regression case, the estimated error term is defined in the same manner; the only difference is in the number of independent variables. The estimated error term for observation i is

(7) $\qquad \hat{u}_i = y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \cdots + \hat{\beta}_k x_{ik})$

where the hat (as before) denotes an estimated value for the parameter. The idea of an estimated error term is the same as in the simple linear case, but in the multiple regression case, there are greater than two dimensions. This makes it difficult to impossible to show the error or deviations graphically. The estimated error term for all observations can be written as follows

(8)
$$\hat{u}_1 = y_1 - (\hat{\beta}_1 + \beta_2 x_{12} + \hat{\beta}_3 x_{13} + \cdots + \hat{\beta}_k x_{1k})$$
$$\hat{u}_2 = y_2 - (\hat{\beta}_1 + \beta_2 x_{22} + \hat{\beta}_3 x_{23} + \cdots + \hat{\beta}_k x_{2k})$$
$$\hat{u}_3 = y_1 - (\hat{\beta}_1 + \beta_2 x_{32} + \hat{\beta}_3 x_{33} + \cdots + \hat{\beta}_k x_{3k}) \ .$$
$$\vdots$$
$$\hat{u}_n = y_n - (\hat{\beta}_1 + \beta_2 x_{n2} + \hat{\beta}_3 x_{n3} + \cdots + \hat{\beta}_k x_{nk})$$

Equation (8) in matrix notation is

(9) $\qquad \hat{U} = Y - \hat{Y} = \hat{Y} - X\hat{\beta}$

where Y and X are as previously defined, $\hat{\beta}$ is the vector of estimated parameters, $\hat{Y}$ is the vector of estimated dependent variables, and $\hat{U}$ is the vector of estimated error terms.

OLS Derivation

It is know time to derive the OLS estimator in matrix form. The objective of the OLS estimator is to minimize the sum of the squared errors. This is no different than the previous simple linear case. The sum of the squared errors or residuals is a scalar, a single number. In matrix form, the estimated sum of squared errors is:

(10)
$$\begin{array}{ccc} SSR & = & \hat{U}' \quad \hat{U} \\ (1 \times 1) & & (1 \times n)\,(n \times 1) \end{array}$$

where the dimensions of each matrix are shown below each matrix and the symbol ' represents the matrix transpose operation. The following example illustrates why this definition is the sum of squares.

Example Sum of Squared Errors Matrix Form. To show in matrix form, the equation d'd is the sum of squares, consider a matrix d of dimension (1 x 3) consisting of the elements 2, 4, 6. Also, recall by taking the transpose, the rows and columns are interchanged. With these elements, the sum of squares equals $2^2 + 4^2 + 6^2 = 56$. In matrix notation, d and d' are

$$d = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \quad d' = \begin{bmatrix} 2 & 4 & 6 \end{bmatrix}$$

The sum of squares equals

$$ss = d'd = \begin{bmatrix} 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 2^2 + 4^2 + 6^2 \end{bmatrix} = 56.$$

This shows multiplying the transpose of a vector by the vector will give the sum of squares of the vector.

With the sum of squared residuals defined and using the definition for calculating the error term, the objective function of OLS can be written as follows:

(11)
$$\min_{\text{w.r.t } \hat{\beta}} \hat{U}'\hat{U} = (Y - X\hat{\beta})'(Y - X\hat{\beta}).$$

In this equation, the only unknowns are the $\hat{\beta}$'s , both the Y and X matrix are know. The Y and X matrix are made up of elements associated with your n-paired observations. Using our knowledge of calculus, we know that if we want to minimize an equation, we can take the first derivative, set the resulting equations equal to zero, and solve for the unknown $\hat{\beta}$'s . To derive the estimator, it is useful to use the following rule of transposing matrices. Using this rule puts equation (11) into a simpler form for derivation. Necessary transpose rule is:

(12) $$(J - LM)' = J' - M'L'$$

where J, L, and M represent matrices conformable for multiplication and addition. Applying transpose rule in equation (12) to equation (11) and then expanding, the following equation is obtained;

(13)
$$\begin{aligned} \min \ \hat{U}'\hat{U} &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= (Y' - \hat{\beta}'X')(Y - X\hat{\beta}) \\ &= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}X'X\hat{\beta} \\ &= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

The last step in simplifying equation (13) relied on the following property $\hat{\beta}'X'Y = Y'X\hat{\beta}$. Note, by matrix multiplication, both sides of this equation results in a scalar. This property is illustrated in the following example.

Example. To show $J'M'L = L'MJ$ holds, define the following matrices:

$$J = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \quad M = \begin{bmatrix} 5 & 6 & 9 \\ 2 & 1 & 7 \end{bmatrix}, \quad L = \begin{bmatrix} 5 \\ 11 \end{bmatrix}.$$

With these matrices, the property is

$$J'M'L = L'MJ$$

$$\begin{bmatrix} 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 6 & 1 \\ 9 & 7 \end{bmatrix} \begin{bmatrix} 5 \\ 11 \end{bmatrix} = \begin{bmatrix} 5 & 11 \end{bmatrix} \begin{bmatrix} 5 & 6 & 9 \\ 2 & 1 & 7 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 3*5+4*6+5*9 & 3*2+4*1+5*7 \end{bmatrix} \begin{bmatrix} 5 \\ 11 \end{bmatrix} = \begin{bmatrix} 5 & 11 \end{bmatrix} \begin{bmatrix} 5*3+6*4+9*5 \\ 2*3+1*4+7*5 \end{bmatrix}$$

$$\begin{bmatrix} 84 & 45 \end{bmatrix} \begin{bmatrix} 5 \\ 11 \end{bmatrix} = \begin{bmatrix} 5 & 11 \end{bmatrix} \begin{bmatrix} 84 \\ 45 \end{bmatrix}$$

$$\begin{bmatrix} 84*5+45*11 \end{bmatrix} = \begin{bmatrix} 5*84+11*45 \end{bmatrix}$$

$$915 = 915$$

We have shown by example the property holds and results in a scalar.

Derivation in Matrix Form

The steps necessary to derive the OLS estimator in matrix form are given in mathematical form in table 1. Each step is described here.

Step 1. This step simply provides the definition of the OLS problem in matrix. That is to minimize the SSR. The above discussion provides the background for this formulation.

Step 2. As we have used throughout this class, taking the partial derivatives and setting them equal zero provides us with candidate points for a minimization or maximization. Here, this step is writing the equation that the partial derivatives will be taken in matrix form.

Step 3. The partial derivatives of the matrix is taken in this step and set equal to zero. Recall, $\hat{\beta}$ is a vector or coefficients or parameters. Because the equation is in matrix form, there are k partial derivatives (one for each parameter in $\hat{\beta}$) set equal to zero. The rules of differentiation are applied to the matrix as follows.
The sum / difference rule is applied to each set of matrices in the equation. Y'Y does not include $\hat{\beta}$, therefore, the partial of Y'Y w.r.t. $\hat{\beta}$ is zero. The second term, $2\hat{\beta}'X'Y$, is a linear term in $\hat{\beta}$. Recall, X'Y is considered a given or constant. Therefore, the derivative of this term is $2X'Y$. The last term, $\hat{\beta}'X'X\hat{\beta}$, is simply a squared term in $\hat{\beta}$

with X'X as constants. The derivative of a squared term is found using the power rule. Applying this rule one obtains $2X'X\hat{\beta}$.

Step 4. Simple matrix algebra is used to rearrange the equation. Recall, the first order conditions are to set the partials equal to zero. First, all terms are divided by the scalar 2. This removes the scalar from the equation. This is simply for ease. Second, $X'Y$ is added to both side of the equation. On the left hand side, the two terms $-X'Y$ and $X'Y$ cancel each other out leaving the null matrix. This step moves $X'Y$ to the right hand side.

Step 5. Finally, $\hat{\beta}$ is found by premultiplying both sides by $(X'X)^{-1}$. Division by matrices is not defined, but multiplying by the inverse is a similar operation. Recall, $(X'X)^{-1}(X'X) = I$, where I is the identity matrix. Multiplying any matrix, A, by I results in A, similar to multiplying by one in linear algebra.

Table 1. Steps Involved in Obtaining the OLS Estimator in Matrix Form

Step	Mathematical Deviation	Step involves
1	$\min \ \hat{U}'\hat{U} = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}X'X\hat{\beta}$	Original problem min. SSR
2	$\partial\hat{U}'\hat{U} \big/ \partial\beta = \partial(Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}) \big/ \partial\beta$	FOC for minimization
3	$\partial\hat{U}'\hat{U} \big/ \partial\beta = 0 - 2X'Y + 2X'X\hat{\beta} = 0$	Use the sum and power rules to take first partial derivative and set equal to zero
4	$X'X\hat{\beta} = X'Y$	Divide both sides by 2 and rearrange by adding X'Y to both sides
5	$\hat{\beta} = (X'X)^{-1}X'Y$	OLS estimator obtained by premultiplying both sides by the inverse of X'X

OLS Estimator Matrix Form

The OLS estimator in matrix form is given by the equation, $\hat{\beta} = (X'X)^{-1}X'Y$. *You must commit this equation to memory and know how to use it.* You will not have to take derivatives of matrices in this class, but know the steps used in deriving the OLS estimator. Notice, the matrix form is much cleaner than the simple linear regression form. More important, the matrix form allows for k unknowns, whereas the simple linear form allowed for only two unknowns an intercept and a slope. We can now estimate more complicated equations. This is important, because most equations that are estimated are not simple linear equations, but rather multiple regressions.

Example. Using the example from the simple linear case, we can show using the matrix form will result in the same OLS estimates. Further, this example shows how the equations are used.

Recall, the example had three paired observations (40, 3), (5, 1), and (10, 2), and the equation we were estimating is $y_i = \beta_1 + \beta_2 x_i + u_i$. In matrix form the equation and observations are:

$$Y = X\beta + U, \quad \text{where}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad Y = \begin{bmatrix} 40 \\ 5 \\ 10 \end{bmatrix}, \quad U = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$$

With these matrices, the OLS estimates for $\hat{\beta}$ are:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

where

$$X' = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix}$$

$$X'X = (\begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 1 & 2 \end{bmatrix})^{-1} = \begin{bmatrix} 1*1+1*1+1*1 & 1*3+1*1+1*2 \\ 3*1+1*1+2*1 & 3*3+1*1+2*2 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

Need to find the det er min ate of $X'X$ to get the inverse

$$|X'X| = 3*14 - 6*6 = 6$$

Next need to find the cofactor and adjo int matrices of $X'X$, note $|\ |$ denote det er min ate.

$$\text{cof}(X'X) = \begin{bmatrix} |14| & -|6| \\ -|6| & |3| \end{bmatrix} = \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$$

$$\text{adj}(X'X) = (\text{cof}(X'X))' = \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$$

$$X'X = \frac{1}{|X'X|}\text{adj}(X'X) = \frac{1}{6}\begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = \begin{bmatrix} \frac{14}{6} & -1 \\ -1 & \frac{1}{2} \end{bmatrix}$$

$$\hat{\beta} = (X'X)^{-1}X'Y = X = \begin{bmatrix} \frac{14}{6} & -1 \\ -1 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 2 \end{bmatrix}\begin{bmatrix} 40 \\ 5 \\ 10 \end{bmatrix} = \begin{bmatrix} \frac{14}{6} & -1 \\ -1 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 55 \\ 145 \end{bmatrix} = \begin{bmatrix} -16.67 \\ 17.5 \end{bmatrix}$$

This is the same result that was obtained for the simple linear regression case. As the number of unknown parameters increase, the only difficulty is the matrix algebra becomes increasingly tedious. Fortunately, computer programs, such as Excel have built in algorithms to estimate equations using OLS. Thus, simplifying the estimation process. However, you still need to understand how the estimator is derived.

SOC

The math necessary to show the SOC in matrix form is beyond the matrix algebra presented in the prerequisites for this class. Similar to the simple linear case, OLS minimizes a squared function. By squaring the error term, the function can only be equal to zero or greater. Intuition tells us such a function will not have a maximum but rather a minimum. We will assume the SOC hold. They will hold using the set up presented here.

Assumptions Made to this Point

As in the simple linear case, very few assumptions have been made to derive the OLS estimator. Because so few assumptions have been made, OLS is a powerful estimation technique. Key point: the two assumptions made are the same as in the simple linear case. We have done nothing new, except expand the methodology to more than two unknown parameters. The two assumptions are 1) the equation to be estimated is linear in the unknown parameters, β, and 2) the FOC could be solved. Neither assumption is particularly restrictive. Again it is important to note, the assumptions say nothing about the statistical distribution of the estimates, just that we can get the estimates. One reason OLS is so powerful is that estimates can be obtained under these fairly unrestrictive assumptions. Because the OLS can be obtained easily, this also results in OLS being misused. The discussion will return to these assumptions and additional assumptions as the OLS estimator is continually derived.

The assumption that the FOC can be solved requires the determinate of X'X to not equal zero. In finding the inverse of X'X, the adjoint matrix of X'X is divided by the determinate of X'X (a scalar). Division by zero is not defined. To foreshadow coming events, this issue of the determinate not equaling zero will be important when we expand on the use of OLS, specifically multicollinearity (upcoming lecture).

Algebraic Properties of the OLS Estimator

Several algebraic properties of the OLS estimator were shown for the simple linear case. As one would expect, these properties hold for the multiple linear case. The properties are simply expanded to include more than one independent variable. The derivation of these properties is not as simple as in the simple linear case. Because of this, the properties are presented, but not derived. The importance of these properties is they are used in deriving goodness-of-fit measures and statistical properties of the OLS estimator.

Algebraic Property 1. The sum of the estimated residuals (error terms) is equal to zero:

(14) $$\sum_{i=1}^{n} \hat{u}_i = 0.$$

The sum of the residuals equally zero implies the mean of the residuals must also equal zero.

Algebraic Property 2. The point $(\overline{Y}, \overline{X})$ will always be on the estimated line. If the mean of each independent variable is used in the estimated equation, the resulting y will equal the mean of the y observations.

Algebraic Property 3. The sample covariance between each individual x_i and the OLS residual $\hat{u}_i$ is equal to zero.

Algebraic Property 4. The mean of the variable, y, will equal the mean of the $\hat{y}$.

KEY POINT: these algebraic properties pertain to scalars; therefore, the number of independent variables does not change these properties. That is, the matrix forms adds nothing to the derivation.

Goodness-of-Fit

The goodness-of-fit measure used for the simple linear case, R^2, is also the measure used in the multiple regression case. Similar to the algebraic properties, because R^2 is a scalar, the matrix form adds nothing to its derivation. Recall, the coefficient of determination, R^2, measures the amount of the sample variation in y that is explained by x. This is given by the equation

$$(15) \qquad R^2 = 1 - \frac{SSR}{SST}.$$

As shown in this equation, the coefficient of determination, R^2, is one minus the ratio of the amount of variation not explained relative to the total variation. Thus, R^2 measures the amount of sample variation in y that is explained by all the independent variables given by the matrix X. R^2 can range from zero (no fit) to one (perfect fit). If X explains no variation in y, the SSE will equal zero. Looking at equation (15), if SSR = SST a value of zero for R^2 is obtained. On the other hand, if X explains all the variation in y, SSR will equal zero. In this case, R^2 equals one. The values of [0 - 1] are just the theoretical range for the coefficient of determination. One will not usually see either of these values when running a regression.

As noted earlier, the coefficient of determination (and its adjusted value discussed later) is the most common measure of the fit of an estimated equation to the observed data. Although, the coefficient of determination is the most common measure, it is not the only measure of the fit of an equation. One needs to look at other measures of fit, that is don't use R^2 as your only gauge of the fit of an estimated equation. Unfortunately, there is not a cutoff value for R^2 that gives a good measure of fit. Further, in economic data it is not uncommon to have low R^2 values. This is a fact of using socio-economic cross-sectional data. We will continue the discussion on R^2 later in this class, when model specification is discussed.

Important Terms / Concepts

n-paired observations
Error Term vs. estimated error term
Residual
Hat symbol

Sum of squares
SSR
SST
SSE
Why OLS is powerful?
Why OLS is misused?
Four Algebraic properties
Goodness-of-fit
R^2 - Coefficient of Determination
Range of R^2
β vs $\hat{\beta}$
Meaning of n, k, i