Basics for Data Analysis

Suduk Kim

Energy Systems Division,
Graduate School,
Ajou University,

Email: suduk@ajou.ac.ac.kr

1 Exercise

$$y = x\beta + \mu \tag{1}$$

The estimator of β, $\hat{\beta}$ is

$$\hat{\beta} = (x'x)^{-1}x'y \tag{2}$$

Suppose y, x, β above are defined as following:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad x = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \tag{3}$$

(HINT): For a matrix A, if we let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then, we know the following holds:

$$\begin{aligned} A^{-1} &= \frac{1}{|A|}(adjA) \\ &= \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \end{aligned} \tag{4}$$

Suppose we actually calculate $(x'x)^{-1}$ and $x'y$ to put them together into $\hat{\beta} = (x'x)^{-1}x'y$,

$$\begin{aligned} x'x &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix}' \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix} \\ &= \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \end{aligned} \tag{5}$$

Using Cramer's rule and from equation (8), we can exercise to get a and b. From equation (4),

$$(x'x)^{-1} = \frac{1}{n\sum x_i^2 - (\sum x_i)^2}\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \tag{6}$$

$$\begin{aligned} x'y &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix}' \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\ &= \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} \end{aligned} \tag{7}$$

Finally from equation (6) and (7),

$$(x'x)^{-1}x'y = \frac{\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}\begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}}{n\sum x_i^2 - (\sum x_i)^2} \tag{8}$$

$$= \frac{\begin{pmatrix} \sum x_i^2 - \sum x_i \sum x_i y_i \\ -\sum x_i \sum y_i + n\sum x_i y_i \end{pmatrix}}{n\sum x_i^2 - (\sum x_i)^2} \tag{9}$$

$$\begin{aligned} b &= \frac{-\sum x_i \sum y_i + n\sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} \\ &= \frac{n\sum x_i y_i - (n\bar{x})(n\bar{y})}{n\sum x_i^2 - (n\bar{x})^2} \\ &= \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \end{aligned} \tag{10}$$

$$\begin{aligned} a &= \frac{n\sum x_i^2 \bar{y} - n\bar{x}\sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} \\ &= \frac{\sum x_i^2 \bar{y} - \bar{x}\sum x_i y_i}{\sum x_i^2 - n\bar{x}^2} \\ &= \frac{(\sum x_i^2 - n\bar{x}^2 + n\bar{x}^2)\bar{y} - \bar{x}\sum x_i y_i}{\sum x_i^2 - n\bar{x}^2} \\ &= \frac{(\sum x_i^2 - n\bar{x}^2)\bar{y} + n\bar{x}^2\bar{y} - \bar{x}\sum x_i y_i}{\sum x_i^2 - n\bar{x}^2} \\ &= \bar{y} - \frac{-n\bar{x}\bar{y} + \sum x_i y_i}{\sum x_i^2 - n\bar{x}^2}\bar{x} \\ &= \bar{y} - b\bar{x} \end{aligned} \tag{11, 12}$$

2 Derivation of OLS(ordinary least squares) Estimator

$$
\begin{aligned}
y &= x\beta + \mu && (13)\\
\hat{y} &= x\hat{\beta} && (14)\\
\hat{\mu} &= y - \hat{y} \\
&= y - x\hat{\beta} && (15)\\
Q &= \hat{\mu}'\hat{\mu} \\
&= (y - x\hat{\beta})'(y - x\hat{\beta}) \\
&= y'y - \hat{\beta}'x'y - y'x\hat{\beta} + \hat{\beta}'x'x\hat{\beta} \\
&= y'y - 2\hat{\beta}'x'y + \hat{\beta}'x'x\hat{\beta} && (16)
\end{aligned}
$$

Let's minimize RSS(residual sum of squares) with respect to $\hat{\beta}$, then,

$$
\begin{aligned}
&\frac{\partial Q}{\partial \beta} = 0 : && (17)\\
\hat{\beta} &= (x'x)^{-1}x'y \\
&= (x'x)^{-1}x'(x\beta + \mu) \\
&= \beta + (x'x)^{-1}x'\mu && (18)\\
E(\hat{\beta}) &= \beta && (19)\\
V(\hat{\beta}) &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \\
&= E\left[(x'x)^{-1}x'\mu\mu'x(x'x)^{-1}\right] \\
&= (x'x)^{-1}x'E(\mu\mu')x(x'x)^{-1} \\
&= \sigma^2(x'x)^{-1} \text{ where } E(\mu\mu') = \sigma^2 I && (20)
\end{aligned}
$$

From equation (13-15), in general, x, y, and β can be represented as,

$$
x = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1k} \\ 1 & x_{22} & \cdots & x_{2k} \\ & & & \vdots \\ 1 & x_{n2} & \cdots & x_{nk} \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \qquad (21)
$$

In this case, we can also get the estimate $\hat{\beta}$ of population parameter β by differentiating the RSS with respect to each $\hat{\beta}_i$. That is, from equation (15), we get

$$
\begin{aligned}
\hat{\mu}'\hat{\mu} &= \sum \mu_i^2 \\
&= \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik})^2 && (22)\\
&\equiv Q
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Q}{\partial \hat{\beta}_1} &= 2\sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik})(-1) = 0 \\
\frac{\partial Q}{\partial \hat{\beta}_2} &= 2\sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik})(-x_{i2}) = 0
\end{aligned}
$$

$$
\begin{aligned}
&\vdots \\
\frac{\partial Q}{\partial \hat{\beta}_k} &= 2\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik})(-x_{ik}) = 0
\end{aligned}
$$

We can rewrite the above equations as following after rearranging them.

$$
\begin{aligned}
n\hat{\beta}_1 + \hat{\beta}_2 \sum x_{i2} + \cdots + \hat{\beta}_k \sum x_{ik} &= \sum y_i \\
\hat{\beta}_1 \sum x_{i2} + \hat{\beta}_2 \sum x_{i2}^2 + \cdots + \hat{\beta}_k \sum x_{i2} x_{ik} &= \sum y_i x_{i2} \\
&\vdots \\
\hat{\beta}_1 \sum x_{ik} + \hat{\beta}_2 \sum x_{ik} x_{i2} + \cdots + \hat{\beta}_k \sum x_{ik}^2 &= \sum y_i x_{ik}
\end{aligned}
$$

$$
\begin{pmatrix}
n & \sum x_{i2} & \cdots & \sum x_{ik} \\
\sum x_{i2} & \sum x_{i2}^2 & \cdots & \sum x_{i2} x_{ik} \\
\vdots & \vdots & \vdots & \vdots \\
\sum x_{ik} & \sum x_{ik} x_{i2} & \cdots & \sum x_{ik}^2
\end{pmatrix}
\begin{pmatrix}
\hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k
\end{pmatrix}
=
\begin{pmatrix}
\sum y_i \\ \sum x_{i2} y_i \\ \vdots \\ \sum x_{ik} y_i
\end{pmatrix}
\tag{23}
$$

We can see that the above equation (23) is in the form of $(x'x)\hat{\beta} = x'y$.

Exercise: Try to obtain $\hat{\beta}$ using the method discussed above when x in equation (13) is

1. a vector of constant 1,
2. a vector of constant 1 and one explanatory variable.

3 BLUE(best linear unbiased estimator)

Best linear unbiased estimator is an estimator which has the least variance among all linear unbiased estimator. Let $(x'x)^{-1}x' = A$,so that $\hat{\beta}$ can be a general form of linear estimator such as

$$\begin{aligned} \hat{\beta} &= (x'x)^{-1}x'y \\ &= Ay \end{aligned} \tag{24}$$

Assume that we have another linear unbiased estimator $\beta^* = C^*y$,

$$\begin{aligned} \beta^* &= C^*y \\ &= (A+C)(x\beta+\mu) \\ &= (A+C)x\beta + (A+C)\mu \\ &= Ax\beta + Cx\beta + (A+C)\mu \\ &= \beta + Cx\beta + (A+C)\mu \end{aligned} \tag{25}$$

then, since β^* is unbiased estimator, $E(\beta^*) = \beta$ is satisfied and it implies $Cx = 0$ should hold from equation (25).

$$\begin{aligned} V(\beta^*) &= E[(A+C)\mu\mu'(A+C)'] \\ &= (A+C)(A+C)'\sigma^2 \end{aligned} \tag{26}$$

Then,

$$\begin{aligned} (A+C)(A+C)' &= AA' + AC' + CA' + CC' \\ &= (x'x)^{-1} + (x'x)^{-1}x'C' + Cx(x'x)^{-1} + CC' \\ &= (x'x)^{-1} + CC' \end{aligned} \tag{27}$$

$$\begin{aligned} V(\beta^*) &= (A+C)(A+C)'\sigma^2 \\ &= \left((x'x)^{-1} + CC'\right)\sigma^2 \\ &= (x'x)^{-1}\sigma^2 + CC'\sigma^2 \\ &= V(\hat{\beta}) + CC'\sigma^2 \geq V(\hat{\beta}) \end{aligned} \tag{28}$$

the above holds since CC' is positive semidefinite.

4 Regarding Distributions

1 Suppose M is symmetric and idempotent matrix $rank(M) = trace(M)$.

2 Suppose $\mu \sim N(0, V)$, then

$$\mu' V^{-1} \mu \sim x^2(k).$$

3 Suppose $\mu \sim N(0, \sigma^2 I_n)$ and M, N are $k \times k$ symmetric and idempotent matrix with $MN = 0$, then $\mu' M \mu$ and $\mu' N \mu$ are stochastically independent.

4 Suppose $\mu \sim x^2(m)$, and $\nu \sim x^2(n)$, then

$$\frac{\mu/m}{\nu/n} \sim F(m, n).$$

5 Suppose $Z \sim N(0,1)$ and $\mu \sim x^2(m)$ are independent, $T = \frac{Z}{\sqrt{\mu/m}}$ is following t-Distribution with degree of freedom m.

$$\begin{aligned} Y &= X\beta + \mu && (29) \\ \hat{Y} &= X\hat{\beta} && (30) \\ \hat{\mu} &= Y - \hat{Y} \\ &= Y - X\hat{\beta} && (31) \\ \hat{\beta} &= \beta + (X'X)^{-1}X'\mu && (32) \end{aligned}$$

If we let $X(X'X)^{-1}X' = M$, $I - X(X'X)^{-1}X' = I - M = N$, then $rank(M) = k$ and $rank(N) = n - k$ holds. Since $\hat{\beta} - \beta = (X'X)^{-1}X'\mu$ From above equation (32),

$$\begin{aligned} X(\hat{\beta} - \beta) &= X(X'X)^{-1}X'\mu = M\mu && (33) \\ \hat{\mu} &= X\beta + \mu - X\hat{\beta} && (34) \\ &= \mu - M\mu = (I - M)\mu = N\mu && (35) \\ \hat{\mu}'\hat{\mu} &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) = \mu' N \mu && (36) \\ \frac{\hat{\mu}'\hat{\mu}}{\sigma^2} &= (\frac{\mu}{\sigma})' N (\frac{\mu}{\sigma}) \sim x^2(n-k) && (37) \\ \frac{(\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta)}{\sigma^2} &= (\frac{\mu}{\sigma})' M (\frac{\mu}{\sigma}) \sim x^2(k) && (38) \end{aligned}$$

5 $t-$Value of $\hat{\beta}_i$, R^2, $\bar{R}^2$, Significance Test of Regression Equation

5.1 $t-$Value of $\hat{\beta}_i$

From $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$, and from $\frac{\hat{\mu}'\hat{\mu}}{\sigma^2} = (\frac{\mu}{\sigma})'N(\frac{\mu}{\sigma}) \sim x^2_{(n-k)}$ we can discuss the method to test the statistical significance of $\hat{\beta}_i$.

First, if we take the estimator of σ as

$$s^2 = \frac{\hat{\mu}'\hat{\mu}}{n-k}$$

, we can show that $E(s^2) = \sigma^2$, since $\frac{(n-k)S^2}{\sigma^2} \sim \chi^2_{(n-k)}$ holds. Therefore, with this following distributional information of $T_{(n-k)}$,

$$\begin{aligned} T_{(n-k)} &= \frac{Z}{\sqrt{\chi^2_{(n-k)}/(n-k)}} && (39) \\ &= \frac{(\hat{\beta}_i - \beta_i)/\sigma\sqrt{(x'x)^{-1}_{ii}}}{\sqrt{\frac{(n-k)s^2}{\sigma^2}/(n-k)}} && (40) \\ &= \frac{\hat{\beta}_i - \beta_i}{s\sqrt{(x'x)^{-1}_{ii}}} && (41) \end{aligned}$$

we can test the null hypothesis, $H_0 : \beta_i = 0$.

5.2 R^2, $\bar{R}^2$

TSS: Total Sum of Squares $Y'Y$
ESS: Explained Sum of Squares $\hat{\beta}'X'X\hat{\beta} = (X\hat{\beta})'(X\hat{\beta})$
RSS: Residual Sum of Squares $\hat{\mu}'\hat{\mu}$
Let $y = Y - \bar{Y}$, then, $y'y = Y'Y - n\bar{Y}^2$.

Exercise: Discuss the reason why $\hat{\beta}'x'\hat{\mu} = \hat{\mu}'x\hat{\beta} = 0$.

$$\begin{aligned} y &= x\hat{\beta} + \hat{\mu} && (42) \\ y'y &= \hat{\beta}'x'x\hat{\beta} + \hat{\mu}'\hat{\mu} && (43) \\ \text{TSS} &= \text{ESS} + \text{RSS} && (44) \\ R^2 &= \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} && (45) \\ &= 1 - \frac{\hat{\mu}'\hat{\mu}}{y'y} && (46) \\ &= 1 - \frac{\hat{\mu}'\hat{\mu}}{Y'Y - n\bar{Y}^2} && (47) \\ \bar{R}^2 &= 1 - \frac{\hat{\mu}'\hat{\mu}/(n-k)}{(y'y - n\bar{Y}^2)/(n-1)} && (48) \\ &= 1 - (1-R^2)\frac{n-k}{n-1} && (49) \end{aligned}$$

5.3 the Significance Test of Regression Equation

Test of null hypothesis $H_0 : \beta_i = 0, i = 2, 3, \cdots, k$.

Let $y = Y - \bar{Y}$, $x = X - \bar{X}$, then

$$\frac{(\hat{\beta} - \beta)' x' x (\hat{\beta} - \beta)}{\sigma^2} = (\frac{\mu}{\sigma})' M (\frac{\mu}{\sigma}) \sim \chi^2_{(k-1)}$$

above will be reduced to

$$\frac{\hat{\beta}' x' x \hat{\beta}}{\sigma^2} \sim \chi^2_{(k-1)}$$

under null hypotheis of H_0. In addition, since $\hat{\beta}' x' x \hat{\beta} = \hat{\beta}' X' Y - n\bar{Y}^2 = y'y - \hat{\mu}'\hat{\mu}$,

$$\frac{(y'y - \hat{\mu}'\hat{\mu})/(k-1)}{(\hat{\mu}'\hat{\mu})/(n-k)} \tag{50}$$

$$= \frac{(\hat{\beta}' X' Y - n\bar{Y}^2)/(k-1)}{(\hat{\mu}'\hat{\mu})/(n-k)} \tag{51}$$

$$= \frac{(\hat{\beta}' X' Y - n\bar{Y}^2)/(k-1)}{(Y'Y - \hat{\beta}' X' Y)/(n-k)} \quad \sim \quad F_(k-1, n-k) \tag{52}$$

6 Hypothesis Testing for the Coefficients Under Constraints

Suppose R is a matrix with $m \times k$, we can set the m constraints on coefficients as $R\beta = r$. Then the constrained estimator Rb follows

$$\begin{aligned} E(Rb) &= r \\ V(Rb) &= E[R(b-\beta)(b-\beta)'R'] = \sigma^2 R(x'x)^{-1}R' \\ (Rb-r)[R(x'x)^{-1}R']^{-1}(Rb-r)/\sigma^2 &\sim \chi^2(m) \\ (Rb-r)[R(x'x)^{-1}R']^{-1}(Rb-r)/(ms^2) &\sim F(m, n-k) \end{aligned}$$

Suppose b is an estimator for β without constraints, and b^* as an estimator for β with constraints, then,

$$\begin{aligned} Y - Xb^* &= (Y - Xb) + (Xb - Xb^*) \\ e^* &= e + X(b - b^*) \\ e^{*'}e^* &= e'e + (b-b^*)'X'X(b-b^*), \ \hat{A}\ddot{u}^{\circ}\acute{i}\text{: e'X} = 0 \\ \frac{(b-b^*)'X'X(b-b^*)/m}{e'e/(n-k)} &= \frac{e^{*'}e^* - e'e}{ms^2} \sim F(m, n-k) \end{aligned}$$

The result of the last equation follows since $(b-b^*)'X'X(b-b^*)/\sigma^2 \sim \chi^2(m)$ holds. Also, the Significance Test of Regression Equation discussed above is a special case of this F-test.

(Proof)
Consider Lagrangean Function to obtain b^*, the estimator for β with constraints,

$$\begin{aligned} L &= \frac{1}{2}(Y - Xb^*)'(Y - Xb^*) + \lambda'(Rb^* - r) \\ \frac{\partial L}{\partial b^*} &= -X'(Y - Xb^*) + R'\lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= Rb^* - r = 0 \\ \lambda &= [R(X'X)^{-1}R]^{-1}R(X'X)^{-1}(X'Y - X'Xb^*) \\ &= [R(X'X)^{-1}R]^{-1}Rb - [R(X'X)^{-1}R]^{-1}Rb^*) \\ &= [R(X'X)^{-1}R]^{-1}(Rb - r) \\ b^* &= b + (X'X)^{-1}R'[R(X'X)^{-1}R]^{-1}(Rb - r) \\ (b-b^*)'X'X(b-b^*) &= (Rb-r)[R(x'x)^{-1}R']^{-1}(Rb-r) \end{aligned}$$

7 Test of Homogeneity of Multiple Regression Function

Suppose there are two samples of our choice with the sample size of n_1, n_2, respectively. Set up two population regression functions, $y_1 = X_1\beta_1 + \mu_1$, and $y_2 = X_2\beta_2 + \mu_2$ from which those samples are supposedly drawn. Then X_1, X_2 here are $n_1 \times k, n_2 \times k$, respectively. Under the null hypothesis of $H_0 : \beta_1 = \beta_2$, the model can be represented as following:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

Null hypothesis with k constraints can be presented as

$$R\beta = (I - I) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = 0$$

. From above discussion, we know that

$$\frac{(e^{*'}e^* - e'e)/k}{e'e/(n_1 + n_2 - 2k)} = F(k, n_1 + n_2 - 2k)$$

follows under null hypothesis. This type of technique is also called as the test of structural break.

8 Prediction

Suppose the null hypothesis above is true (or there is no structural break). Make a prediction on y_2 based on the estimator of β_1, b_1. In this case, when X_2 is provided as explanatory variable for y_2, then predicted value of y_2 is $X_2 b_1 = X_2(X_1'X_1)^{-1}X_1'y_1$, and $E(X_2 b_1) = E(y_2)$ will hold. Prediction error (d) is

$$\begin{aligned} d &= y_2 - X_2 b_1 = X_2\beta_2 + \mu_2 - X_2\beta_1 - X_2(X_1'X_1)^{-1}X_1'\mu_1 \\ E(d) &= X_2\beta_2 - X_2\beta_1 = 0 \\ V(d) &= V(\mu_2 - X_2(X_1'X_1)^{-1}X_1'\mu_1) \\ &= E(\mu_2 - X_2(X_1'X_1)^{-1}X_1'\mu_1)(\mu_2 - X_2(X_1'X_1)^{-1}X_1'\mu_1)' \\ &= \sigma^2[I + X_2(X_1'X_1)^{-1}X_2'] \end{aligned}$$

Since we know $(y_2 - X_2 b_1)'[I + X_2(X_1'X_1)^{-1}X_2']^{-1}(y_2 - X_2 b_1)/\sigma^2 \sim \chi^2(n_2)$, $\frac{y_2 - X_2 b_1}{[I + X_2(X_1'X_1)^{-1}X_2']^{1/2} s_1} \sim t_(n-k)$, and especially when $n_2 = 1$. From this, interval estimation for y_2 can be conducted.

9 Example 1

Following example is for the estimation of consumption function. It is noted that one of explanatory variable is a time trend.

Understanding the Regression Results:

- Constant term usually do not have any meaning. Test the size of R^2 for the regression with constant only.
- What is the meaning of coefficient?
- Try to understand the meaning of $R^2, \bar{R}^2, F_{2,12}, F_{1,13}$.
- Consider whether it is desirable to include the variable 'Time'. (Hint:t-test and F-test.)

Regression Results Under No Constraint

$$\begin{array}{rcccccc} \hat{Y}_i & = & 53.16 & + & 0.72X_{i2} & + & 2.73X_{i3} \\ & & (13.02) & & (0.05) & & (0.85) \\ t & = & (4.08) & & (14.91) & & (3.22) \\ p\text{-value} & = & (0.001) & & (0.000) & & (0.003) \\ df & = & 12 & & R^2 = 0.9988 & & F_{2,12} = 5128.88 \\ & & & & \bar{R}^2 = 0.9986 & & RSS_{ur} = 77.1692 \end{array} \tag{53}$$

Regression Results Under Constraints

$$\begin{array}{rcccccc} \hat{Y}_i & = & 12.76 & + & 0.88X_{i2} & & \\ & & (4.68) & & (0.011) & & \\ t & = & (2.73) & & (77.12) & & \\ p\text{-value} & = & (0.017) & & (0.000) & & \\ df & = & 13 & & R^2 = 0.9978 & & F_{1,13} = 5947.72 \\ & & & & \bar{R}^2 = 0.9976 & & RSS_r = 144.0347 \end{array} \tag{54}$$

$$\begin{aligned} F_(m, n-k) &= \frac{(e^{*'}e^* - e'e)/m}{e'e/(n-k)} => \frac{(RSS_r - RSS_{ur})/m}{RSS_{ur}/(n-k)} \\ &= \frac{(R^2 - R^{*2})/m}{(1-R^2)/(n-k)} \\ => & \frac{(144.0347 - 77.1692)/1}{77.1692/12} = 10.3978 \\ => & \frac{(0.9988 - 0.9978)/1}{(1-0.9988)/12} = 10.3978 \end{aligned}$$

** $F_{1,12,0.05} = 4.75$, $t^2 = (3.2246)^2 = 10.3978$. p-value $= 0.0073$

Table 1: Personal Consumption Expenditure(PCE) and Personal Disposable Income (PDI), US 1956-70. (Unit: bill. 1958 dollars)

PCE (Y)	PDI (X_2)	Time (X_3)
281.4	309.3	1956 =1
288.1	316.1	1957=2
290.0	318.8	1958=3
307.3	333.0	1959=4
316.1	340.3	1960=5
322.5	350.5	1961=6
338.4	367.2	1962=7
353.3	381.2	1963=8
373.7	408.1	1964=9
397.7	434.8	1965=10
418.1	458.9	1966=11
430.1	477.5	1967=12
452.7	499.0	1968=13
469.1	513.5	1969=14
476.9	533.2	1970=15

10 Exercise 2

Using Cobb-Douglas production function, we can test capital and labor productivity. We can also test the constraints on the coefficients. (For example, if the sum of capital and labor productivity equals 1 in Cobb-Douglas production function, it means constant returns to scale.)

If $Y = AK^{\alpha}L^{\beta}$, then, the production function we are going to estimate after taking ln on both sides takes the form of

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{i2} + \beta_3 \ln X_{i3} + \mu_i.$$

where $\beta_1 = \ln A, \beta_2 = \alpha, \beta_3 = \beta, \ln X_{i2} = \ln K, \ln X_{i3} = \ln L$ÀÌ´Ù.

To test the null hypothesis $H_0 : \beta_2 + \beta_3 = 1$, regression function under constraints would be

$$\begin{aligned} \ln Y_i &= \beta_1 + (1-\beta_3)\ln X_{i2} + \beta_3 \ln X_{i3} + \mu_i \\ \ln Y_i - \ln X_{i2} &= \beta_1 + \beta_3(\ln X_{i3} - \ln X_{i2}) + \mu_i \\ \ln(Y_i/\ln X_{i2}) &= \beta_1 + \beta_3 \ln(X_{i3}/\ln X_{i2}) + \mu_i \end{aligned}$$

Regression Results Under No Constraint

$$\begin{array}{rcccccc} \widehat{\ln Y_i} & = & -3.34 & + & 1.50 \ln X_{2i} & + & 0.49 \ln X_{3i} \\ & & (2.45) & & (0.54) & & (0.10) \\ t & = & (-1.36) & & (2.78) & & (4.80) \\ df & = & 12 & & R^2 = 0.8890 & & F_{2,12} = 48.068 \\ & & & & \bar{R}^2 = 0.8705 & & RSS_{ur} = 0.067 \end{array} \tag{55}$$

Table 2: Real GDP, Labor Input, Capital Input of Agricultural Sector, Taiwan, 1958-1972.

Year	Real GDP	Labor Input	Capital Input
1958	16607.7	275.5	17803.7
1959	17511.3	274.4	18096.8
1960	20171.2	269.7	18271.8
1961	20932.9	267.0	19167.3
1962	20406.0	267.8	19647.6
1963	20831.6	275.0	20803.5
1964	24806.3	283.0	22076.6
1965	26465.8	300.7	23445.2
1966	27403.0	307.5	24939.0
1967	28628.7	303.7	26713.7
1968	29904.5	304.7	29957.8
1969	27508.2	298.6	31585.9
1970	29035.5	295.5	33475.5
1971	29281.5	299.0	34821.8
1972	31535.8	288.1	41794.3

Regression Results Under Constraints

$$\begin{array}{rcccc} \ln \hat{Y_i}/X_{i2} & = & 1.71 & + & 0.61 \ln(X_{3i}/X_{2i}) \\ & & (0.42) & & (0.09) \\ t & = & (4.17) & & (6.57) \\ df & = & 13 & R^2 = 0.7685 & F_{1,13} = 43.161 \\ & & & \bar{R}^2 = 0.7507 & RSS_r = 0.091 \end{array} \tag{56}$$

$$F_(m, n-k) = \frac{(e^{*'}e^* - e'e)/m}{e'e/(n-k)} => \frac{(RSS_r - RSS_{ur})/1}{RSS_{ur}/12} = F_(1, 12)$$

$F = \frac{(0.8890-0.7685)/1}{(1-0.8890)/12} = 4.3587$. $F_{1,12,0.05} = 4.75$.

(**Homework)** Obtain the test statistics of F using RSS_r, RSS_{ur}.

11 Exercise 3

Estimation of Chicken Demand Function for US(1960-1982).

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \beta_4 \ln X_{4i} + \beta_5 \ln X_{5i} + \mu_i.$$

From the demand theory of economics, we know the sign of $\beta_2 > 0, \beta_3 < 0$, β_4, β_5 are different depending on whether these are for independent good, substitutes, or complements.

The null hypothesis $H_0 : \beta_4 = \beta_5 = 0$ tells that chicken demand is not affected by the price of pork and beef. That is, chicken is independent good of pork and beef. Then under this constraint, the regression function will be

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \mu_i.$$

Regression Results Under Constraints

$$\begin{array}{rcccc} \widehat{\ln Y_i} & = & 2.03 & + & 0.45 \ln X_{2i} \qquad -0.38 \ln X_{3i} \\ & & (0.12) & & (0.02) \qquad\quad (0.06) \\ df & = & 20 & & R^2 = 0.9801 \qquad F_{2,20} = \\ & & & & \bar{R}^2 = 0.7507 \qquad RSS_r = \end{array} \tag{57}$$

Regression Results Under No Constraint

$$\begin{array}{rcccccccc} \widehat{\ln Y_i} & = & 2.19 & + & 0.34 \ln X_{2i} & -0.50 \ln X_{3i} & + & 0.15 \ln X_{4i} & + & 0.09 \ln X_{5i} \\ & & (0.16) & & (0.08) & (0.11) & & (0.10) & & (0.10) \\ df & = & 18 & & R^2 = 0.9823 & F_{4,18} = & & & & \\ & & & & \bar{R}^2 = 0.8705 & RSS_{ur} = & & & & \end{array} \tag{58}$$

$$F_(m, n-k) = \frac{(e^{*'} e^* - e'e)/m}{e'e/(n-k)}$$

(**Homework)** Obtain the test statistics F using RSS_r, RSS_{ur}. Check to see if $F = 1.124$. $F_{2,18,0.05} = 3.55$ for your reference.

Table 3: Chicken Demand of US 1960-1982

Year	Y	X_2	X_3	X_4	X_5	X_6
1960	27.8	397.5	42.2	50.7	78.3	65.8
1961	29.9	413.3	38.1	52.0	79.2	66.9
1962	29.8	439.2	40.3	54.0	79.2	67.8
1963	30.8	459.7	39.5	55.3	79.2	69.6
1964	31.2	492.9	37.3	54.7	77.4	68.7
1965	33.3	528.6	38.1	63.7	80.2	73.6
1966	35.6	560.3	39.3	69.8	80.4	76.3
1967	36.4	624.6	37.8	65.9	83.9	77.2
1968	36.7	666.4	38.4	64.5	85.5	78.1
1969	38.4	717.8	40.1	70.0	93.7	84.7
1970	40.4	768.2	38.6	73.2	106.1	93.3
1971	40.3	843.3	39.8	67.8	104.8	89.7
1972	41.8	911.6	39.7	79.1	114.0	100.7
1973	40.4	931.1	52.1	95.4	124.1	113.5
1974	40.7	1021.5	48.9	94.2	127.6	115.3
1975	40.1	1165.9	58.3	123.5	142.9	136.7
1976	42.7	1349.6	57.9	129.9	143.6	139.2
1977	44.1	1449.4	56.5	117.6	139.2	132.0
1978	46.7	1575.5	63.7	130.9	165.5	132.1
1979	50.6	1759.1	61.6	129.8	203.3	154.4
1980	50.1	1994.2	58.9	128.0	219.6	174.9
1981	51.7	2258.1	66.4	141.0	221.6	180.8
1982	52.9	2478.7	70.4	168.2	232.6	189.4

Y: Chicken Consumption Per Capita (Unit:lb)
X_2: Real Disposable Income per Capita (Unit: \$)
X_3: Real Retail Price of Chicken (Unit:lb, cent)
X_4: Real Retail Price of Pork (Unit:lb, cent)
X_5: Real Retail Price of Beef (Unit:lb, cent)
X_6: Substitute Price Index (Weighted Average of real retail price of Pork and beef)

Table 4: Private Saving and Income, Britain,1946-1963 (Unit: Mil. Pound)

Year	Savings	Income
1946	0.36	8.8
1947	0.21	9.4
1948	0.08	10.0
1959	0.20	10.6
1950	0.10	11.0
1951	0.12	11.9
1952	0.41	12.7
1953	0.50	13.5
1954	0.43	14.3
1955	0.59	15.5
1956	0.90	16.7
1957	0.95	17.7
1958	0.82	18.6
1969	1.04	19.7
1960	1.53	21.1
1961	1.94	22.8
1962	1.75	23.9
1963	1.99	25.2