5 Restricted OLS (制約付き最小二乗法)

1. Let $\tilde{\beta}$ be the restricted estimator.

   Consider the linear restriction: $R\beta = r$.

2. Minimize $(y - X\tilde{\beta})'(y - X\tilde{\beta})$ subject to $R\tilde{\beta} = r$.

   Let L be the Lagrangian for the minimization problem.

$$L = (y - X\tilde{\beta})'(y - X\tilde{\beta}) - 2\tilde{\lambda}'(R\tilde{\beta} - r)$$

Because $\tilde{\beta}$ and $\tilde{\lambda}$ minimize the Lagrangian L,

$$\frac{\partial L}{\partial \tilde{\beta}} = -2X'(y - X\tilde{\beta}) - 2R'\tilde{\lambda} = 0$$

$$\frac{\partial L}{\partial \tilde{\lambda}} = -2(R\tilde{\beta} - r) = 0.$$

(*) Remember that $\dfrac{\partial a'x}{\partial x} = a$ and $\dfrac{\partial x'Ax}{\partial x} = (A + A')x$.

From $\dfrac{\partial L}{\partial \tilde{\beta}} = 0$, we obtain:

$$\tilde{\beta} = (X'X)^{-1}X'y + (X'X)^{-1}R'\tilde{\lambda} = \hat{\beta} + (X'X)^{-1}R'\tilde{\lambda}.$$

Multiplying R from the left, we have:

$$R\tilde{\beta} = R\hat{\beta} + R(X'X)^{-1}R'\tilde{\lambda}.$$

Because $R\tilde{\beta} = r$ has to be satisfied, we have the following expression:

$$r = R\hat{\beta} + R(X'X)^{-1}R'\tilde{\lambda}.$$

Therefore, solving the above equation with respect to $\tilde{\lambda}$, we obtain:

$$\tilde{\lambda} = \Big(R(X'X)^{-1}R'\Big)^{-1}(r - R\hat{\beta})$$

Substituting $\tilde{\lambda}$ into $\tilde{\beta} = \hat{\beta} + (X'X)^{-1}R'\tilde{\lambda}$, the restricted OLSE is given by:

$$\tilde{\beta} = \hat{\beta} + (X'X)^{-1}R'\Big(R(X'X)^{-1}R'\Big)^{-1}(r - R\hat{\beta}).$$

(a) The expectation of $\tilde{\beta}$ is:

$$\begin{aligned}
\mathrm{E}(\tilde{\beta}) &= \mathrm{E}(\hat{\beta}) + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\mathrm{E}(\hat{\beta})) \\
&= \beta + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
&= \beta,
\end{aligned}$$

because of $R\beta = r$.

Thus, it is shown that $\tilde{\beta}$ is unbiased.

(b) The variance of $\tilde{\beta}$ is as follows.

First, rewrite as follows:

$$\begin{aligned}
(\tilde{\beta}-\beta) &= (\hat{\beta}-\beta) + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(R\beta - R\hat{\beta}) \\
&= (\hat{\beta}-\beta) - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - R\beta) \\
&= (\hat{\beta}-\beta) - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(\hat{\beta}-\beta) \\
&= \left(I_k - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R\right)(\hat{\beta}-\beta) \\
&= W(\hat{\beta}-\beta),
\end{aligned}$$

where $W \equiv I_k - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R$.

Then, we obtain the following variance:

$$
\begin{aligned}
\mathrm{V}(\tilde{\beta}) &\equiv \mathrm{E}((\tilde{\beta}-\beta)(\tilde{\beta}-\beta)') = \mathrm{E}(W(\hat{\beta}-\beta)(\hat{\beta}-\beta)'W') \\
&= W\mathrm{E}((\hat{\beta}-\beta)(\hat{\beta}-\beta)')W' = W\mathrm{V}(\hat{\beta})W' = \sigma^2 W(X'X)^{-1}W' \\
&= \sigma^2\Big(I-(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R\Big)(X'X)^{-1} \\
&\qquad\qquad\qquad \times\Big(I-(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R\Big)' \\
&= \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R(X'X)^{-1} \\
&= \mathrm{V}(\hat{\beta}) - \sigma^2(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)^{-1}R(X'X)^{-1}
\end{aligned}
$$

Thus, $\mathrm{V}(\hat{\beta}) - \mathrm{V}(\tilde{\beta})$ is positive definite.

3. Another solution:

Again, write the first-order condition for minimization:

$$\frac{\partial L}{\partial \tilde{\beta}} = -2X'(y - X\tilde{\beta}) - 2R'\tilde{\lambda} = 0,$$
$$\frac{\partial L}{\partial \tilde{\lambda}} = -2(R\tilde{\beta} - r) = 0,$$

which can be written as:

$$X'X\tilde{\beta} - R'\tilde{\lambda} = X'y,$$
$$R\tilde{\beta} = r.$$

Using the matrix form:

$$\begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ -\tilde{\lambda} \end{pmatrix} = \begin{pmatrix} X'y \\ r \end{pmatrix}.$$

The solutions of $\tilde{\beta}$ and $-\tilde{\lambda}$ are given by:

$$\begin{pmatrix} \tilde{\beta} \\ -\tilde{\lambda} \end{pmatrix} = \begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ r \end{pmatrix}.$$

(*) Formula to the inverse matrix:

$$\begin{pmatrix} A & B \\ B' & D \end{pmatrix}^{-1} = \begin{pmatrix} E & F \\ F' & G \end{pmatrix},$$

where E, F and G are given by:

$$E = (A - BD^{-1}B')^{-1} = A^{-1} + A^{-1}B(D - B'A^{-1}B)^{-1}B'A^{-1}$$

$$F = -(A - BD^{-1}B')^{-1}BD^{-1} = -A^{-1}B(D - B'A^{-1}B)^{-1}$$

$$G = (D - B'A^{-1}B)^{-1} = D^{-1} + D^{-1}B'(A - BD^{-1}B')^{-1}BD^{-1}$$

In this case, E and F correspond to:

$$E = (X'X)^{-1} - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}$$

$$F = (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}.$$

Therefore, $\tilde{\beta}$ is derived as follows:

$$\begin{aligned}\tilde{\beta} &= EX'y + Fr \\ &= \hat{\beta} + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(r - R\hat{\beta}).\end{aligned}$$

The variance is:

$$\mathrm{V}\begin{pmatrix} \tilde{\beta} \\ -\tilde{\lambda} \end{pmatrix} = \sigma^2 \begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}^{-1}.$$

Therefore, $\mathrm{V}(\tilde{\beta})$ is:

$$\mathrm{V}(\tilde{\beta}) = \sigma^2 E = \sigma^2\Big((X'X)^{-1} - (X'X)^{-1}R'\Big(R(X'X)^{-1}R'\Big)^{-1}R(X'X)^{-1}\Big)$$

Under the restriction: $R\beta = r$,

$$\mathrm{V}(\hat{\beta}) - \mathrm{V}(\tilde{\beta}) = \sigma^2(X'X)^{-1}R'\Big(R(X'X)^{-1}R'\Big)^{-1}R(X'X)^{-1}$$

is positive definite.

6 F Distribution (Restricted and Unrestricted OLSs)

1. As mentioned above, under the null hypothesis $H_0 : R\beta = r$,

$$\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/G}{(y - X\hat{\beta})'(y - X\hat{\beta})/(n - k)} \sim F(G, n - k),$$

where $G = \text{Rank}(R)$.

Using $\tilde{\beta} = \hat{\beta} + (X'X)^{-1}R'\Big(R(X'X)^{-1}R'\Big)^{-1}(r - R\hat{\beta})$, the numerator is rewritten as follows:

$$(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) = (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}).$$

Moreover, the denominator is rewritten as follows:

$$
\begin{aligned}
(y - X\tilde{\beta})'(y - X\tilde{\beta}) =& (y - X\hat{\beta} - X(\tilde{\beta} - \hat{\beta}))'(y - X\hat{\beta} - X(\tilde{\beta} - \hat{\beta})) \\
=& (y - X\hat{\beta})'(y - X\hat{\beta}) + (\tilde{\beta} - \hat{\beta})'X'X(\tilde{\beta} - \hat{\beta}) \\
& -(y - X\hat{\beta})'X(\tilde{\beta} - \hat{\beta}) - (\tilde{\beta} - \hat{\beta})'X'(y - X\hat{\beta}) \\
=& (y - X\hat{\beta})'(y - X\hat{\beta}) + (\tilde{\beta} - \hat{\beta})'X'X(\tilde{\beta} - \hat{\beta}).
\end{aligned}
$$

$X'(y - X\hat{\beta}) = X'e = 0$ is utilized.

Summarizing, we have following representation:

$$
\begin{aligned}
(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r) &= (\tilde{\beta} - \hat{\beta})'X'X(\tilde{\beta} - \hat{\beta}) \\
&= (y - X\tilde{\beta})'(y - X\tilde{\beta}) - (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= \tilde{u}'\tilde{u} - e'e,
\end{aligned}
$$

where e and $\tilde{u}$ are the restricted residual and the unrestricted residual, i.e., $e = y - X\hat{\beta}$ and $\tilde{u} = y - X\tilde{\beta}$.

Therefore, we obtain the following result:

$$
\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/G}{(y - X\hat{\beta})'(y - X\hat{\beta})/(n - k)} = \frac{(\tilde{u}'\tilde{u} - e'e)/G}{e'e/(n - k)} \sim F(G, n - k).
$$

7 Example: F Distribution (Restricted OLS and Unrestricted OLS)

Date file $\Longrightarrow$ cons99.txt (Next slide)

Each column denotes year, nominal household expenditures (家計消費，10 billion yen), household disposable income (家計可処分所得，10 billion yen) and household expenditure deflator (家計消費デフレータ，1990=100) from the left.

1955	5430.1	6135.0	18.1	1970	37784.1	45913.2	35.2	1985	185335.1	220655.6	93.9
1956	5974.2	6828.4	18.3	1971	42571.6	51944.3	37.5	1986	193069.6	229938.8	94.8
1957	6686.3	7619.5	19.0	1972	49124.1	60245.4	39.7	1987	202072.8	235924.0	95.3
1958	7169.7	8153.3	19.1	1973	59366.1	74924.8	44.1	1988	212939.9	247159.7	95.8
1959	8019.3	9274.3	19.7	1974	71782.1	93833.2	53.3	1989	227122.2	263940.5	97.7
1960	9234.9	10776.5	20.5	1975	83591.1	108712.8	59.4	1990	243035.7	280133.0	100.0
1961	10836.2	12869.4	21.8	1976	94443.7	123540.9	65.2	1991	255531.8	297512.9	102.5
1962	12430.8	14701.4	23.2	1977	105397.8	135318.4	70.1	1992	265701.6	309256.6	104.5
1963	14506.6	17042.7	24.9	1978	115960.3	147244.2	73.5	1993	272075.3	317021.6	105.9
1964	16674.9	19709.9	26.0	1979	127600.9	157071.1	76.0	1994	279538.7	325655.7	106.7
1965	18820.5	22337.4	27.8	1980	138585.0	169931.5	81.6	1995	283245.4	331967.5	106.2
1966	21680.6	25514.5	29.0	1981	147103.4	181349.2	85.4	1996	291458.5	340619.1	106.0
1967	24914.0	29012.6	30.1	1982	157994.0	190611.5	87.7	1997	298475.2	345522.7	107.3
1968	28452.7	34233.6	31.6	1983	166631.6	199587.8	89.5				
1969	32705.2	39486.3	32.9	1984	175383.4	209451.9	91.8				

Estimate using TSP 5.0.

```
LINE  **********************************************
|         freq a;
|         smpl 1955 1997;
|         read(file='cons99.txt') year cons yd price;
|         rcons=cons/(price/100);
|         ryd=yd/(price/100);
|         d1=0.0;
|         smpl 1974 1997;
|         d1=1.0;
|         smpl 1956 1997;
|         d1ryd=d1*ryd;
|         olsq rcons c ryd;
|         olsq rcons c d1 ryd d1ryd;
|         end;
****************************************************
```

```
                  Equation   1
                  ============

  Method of estimation = Ordinary Least Squares


Dependent variable: RCONS
Current sample:  1956 to 1997
Number of observations:  42

 Mean of dependent variable = 149038.
Std. dev. of dependent var. = 78147.9
   Sum of squared residuals = .127951E+10
      Variance of residuals = .319878E+08
   Std. error of regression = 5655.77
                  R-squared = .994890
```

```
        Adjusted R-squared = .994762
   Durbin-Watson statistic = .116873
  F-statistic (zero slopes) = 7787.70
 Schwarz Bayes. Info. Crit. = 17.4101
 Log of likelihood function = -421.469

          Estimated     Standard
Variable  Coefficient     Error         t-statistic
C         -3317.80      1934.49         -1.71508
RYD       .854577       .968382E-02     88.2480
```

```
                Equation   2
                ============

  Method of estimation = Ordinary Least Squares


 Dependent variable: RCONS
 Current sample:  1956 to 1997
 Number of observations:  42

  Mean of dependent variable = 149038.
 Std. dev. of dependent var. = 78147.9
    Sum of squared residuals = .244501E+09
       Variance of residuals = .643423E+07
    Std. error of regression = 2536.58
                   R-squared = .999024
```

Adjusted R-squared = .998946
Durbin-Watson statistic = .420979
F-statistic (zero slopes) = 12959.1
Schwarz Bayes. Info. Crit. = 15.9330
Log of likelihood function = -386.714

Variable	Estimated Coefficient	Standard Error	t-statistic
C	4204.11	1440.45	2.91861
D1	-39915.3	3154.24	-12.6545
RYD	.786609	.015024	52.3561
D1RYD	.194495	.018731	10.3839

1. `Equation 1`

Significance test:

`Equation 1` is:

$$\texttt{RCONS} = \beta_1 + \beta_2 \texttt{RYD}$$

$H_0:\ \beta_2 = 0$

(No.1) t Test $\implies$ Compare 88.2480 and $t(42-2)$.

(No.2) F Test $\implies$ Compare $\dfrac{R^2/G}{(1-R^2)/(n-k)} = \dfrac{.994890/1}{(1-.994890)/(42-2)} =$ 7787.8 and $F(1,40)$. Note that $\sqrt{7787.8} = 88.2485$.

1% point of $F(1, 40) = 7.31$

$H_0:\ \beta_2 = 0$ is rejected.

2. `Equation 2:`

$$\texttt{RCONS} = \beta_1 + \beta_2\texttt{D1} + \beta_3\texttt{RYD} + \beta_4\texttt{RYD} \times \texttt{D1}$$

$H_0:\ \beta_2 = \beta_3 = \beta_4 = 0$

F Test $\Longrightarrow$ Compare $\dfrac{R^2/G}{(1-R^2)/(n-k)} = \dfrac{.999024/3}{(1-.999024)/(42-4)} = 12965.5$ and $F(3, 38)$.

1% point of $F(3, 38) = 4.34$

$H_0: \beta_2 = \beta_3 = \beta_4 = 0$ is rejected.

3. `Equation 1` vs. `Equation 2`

   Test the structural change between 1973 and 1974.

   `Equation 2` is:

$$\mathtt{RCONS} = \beta_1 + \beta_2 \mathtt{D1} + \beta_3 \mathtt{RYD} + \beta_4 \mathtt{RYD} \times \mathtt{D1}$$

   $H_0: \beta_2 = \beta_4 = 0$

Restricted OLS $\Longrightarrow$ `Equation 1`

Unrestricted OLS $\Longrightarrow$ `Equation 2`

$$\frac{(\tilde{u}'\tilde{u} - e'e)/G}{e'e/(n-k)} = \frac{(.127951\mathtt{E}+10 - .244501\mathtt{E}+09)/2}{.244501\mathtt{E}+09/(42-4)} = 80.43$$

which should be compared with $F(2, 38)$.

1% point of $F(2, 38) = 5.211 < 80.43$

$H_0:\ \beta_2 = \beta_4 = 0$ is rejected.

$\Longrightarrow$ The structure was changed in 1974.

8 Generalized Least Squares Method (GLS, 一般化最小自乗法)

1. Regression model: $y = X\beta + u, \qquad u \sim N(0, \sigma^2\Omega)$

2. **Heteroscedasticity (不等分散，不均一分散)**

$$\sigma^2\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{pmatrix}$$

First-Order Autocorrelation (一階の自己相関，系列相関)

In the case of time series data, the subscript is conventionally given by t, not i .

$$u_t = \rho u_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2)$$

$$\sigma^2\Omega = \frac{\sigma_\epsilon^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

$$V(u_t) = \sigma^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}$$

3. The Generalized Least Squares (GLS，一般化最小二乘法) estimator of β, denoted by b, solves the following minimization problem:

$$\min_b \; (y - Xb)'\Omega^{-1}(y - Xb)$$

The GLSE of β is:

$$b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$