A Generic Linear Regression Equation

Scalar Formulation of the Population Regression Equation (PRE)

- **With observation subscripts**. For the i-th observation, the **scalar formulation of the PRE** is written as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \qquad \forall\ i$$

$$= \beta_0 + \sum_{j=1}^{j=k} \beta_j X_{ji} + u_i \qquad \forall\ i$$

$$= \sum_{j=0}^{j=k} \beta_j X_{ji}\ u_i\ , \qquad X_{0i} = 1\ \forall\ i \qquad \forall\ i$$

where

Y_i ≡ the i-th population value of the regressand (*observable* variable);

X_{ji} ≡ the i-th population value of the j-th regressor, j = 0, 1, 2, …, k (*observable* variables);

β_j ≡ the partial regression coefficient of X_{ij}, j = 0, 1, 2, …, k (*unknown* constants);

u_i ≡ the i-th population value of the *unobservable* random error term (*unobservable* variable).

Note:

(1) **Lower case "k"** denotes the **number of *slope* coefficients** in the PRF.

(2) **Upper case "K"** denotes the ***total* number of regression coefficients** in the PRF.

(3) Therefore: **K = k + 1**.

OLS Estimation of the PRE: the OLS Sample Regression Equation (SRE)

- **The Sample Data:** The **sample data** consist of **N observations** on the regressand Y and the regressors $X_1, X_2, \cdots X_k$, which together comprise the observable variables in the population regression equation.

  Each of these N sample observations can be written as a row vector:

  $$(Y_i \quad X_{1i} \quad X_{2i} \quad \cdots \quad X_{ki}) \qquad i = 1, \ldots, N$$

- **OLS Estimation of the PRE:** yields the **OLS sample regression equation** – the **OLS SRE**:

  $$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i \qquad i = 1, \ldots, N \qquad \text{(OLS SRE)}$$

  where

  $\hat{\beta}_j$ = the OLS estimate of the regression coefficient β_j (j = 0, 1, …, k);

  $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}$ = the i-th OLS residual (i = 1, …, N)

A First Guide to Hypothesis Testing in Linear Regression Models

1. Tests of *One* Coefficient Restriction: *One* Restriction on *One* Coefficient

H_0 specifies ***only one equality*** **restriction** on ***one coefficient***.

❑ *Two-Tail* Tests of *One* Restriction on *One* Coefficient

- ***Example:*** H_0: $\beta_j = b_j$ versus H_1: $\beta_j \neq b_j$ where b_j is a specified constant.
- ***Use:*** either a ***two-tail*** **t-test** or an **F-test**.
- ***Test Statistics:***

⇨ $t(\hat{\beta}_j) = \dfrac{\hat{\beta}_j - \beta_j}{\hat{se}(\hat{\beta}_j)} \sim t[N-K]$ ⇒ ***sample value*** $= t_0(\hat{\beta}_j) = \dfrac{\hat{\beta}_j - b_j}{\hat{se}(\hat{\beta}_j)}$.

⇨ $F(\hat{\beta}_j) = \dfrac{\left(\hat{\beta}_j - \beta_j\right)^2}{V\hat{a}r(\hat{\beta}_j)} \sim F[1, N-K]$ ⇒ ***sample value*** $= F_0(\hat{\beta}_j) = \dfrac{\left(\hat{\beta}_j - b_j\right)^2}{V\hat{a}r(\hat{\beta}_j)}$.

Note: $\left[t(\hat{\beta}_j)\right]^2 = F(\hat{\beta}_j)$ or $t(\hat{\beta}_j) = \sqrt{F(\hat{\beta}_j)}$ and $\left[t_{\alpha/2}[N-K]\right]^2 = F_\alpha[1, N-K]$.

- ***Decision Rules – Two-Tail Tests of One Coefficient Restriction on One Coefficient:***

Reject* H₀** if $|t_0| > t_{\alpha/2}[N-K]$ *or* ***two-tail p-value for $t_0 = \Pr(|t| > |t_0|) < \alpha$;
$F_0 > F_\alpha[1, N-K]$ *or* ***p-value*** for $F_0 = \Pr(F > F_0) < \alpha$.

Retain* H₀** if $|t_0| \le t_{\alpha/2}[N-K]$ *or* ***two-tail p-value for $t_0 = \Pr(|t| > |t_0|) \ge \alpha$;
$F_0 \le F_\alpha[1, N-K]$ *or* ***p-value*** for $F_0 = \Pr(F > F_0) \ge \alpha$.

❑ ***One-Tail*** **Tests of *One* Restriction on *One* Coefficient**

- ***Examples:*** H_0: $\beta_j = b_j$ (or $\beta_j \ge b_j$) versus H_1: $\beta_j < b_j$ a ***left-tail*** **test**
  H_0: $\beta_j = b_j$ (or $\beta_j \le b_j$) versus H_1: $\beta_j > b_j$ a ***right-tail*** **test**

- ***Use:*** a ***one-tail*** **t-test**.

- ***Test Statistic:***

⇨ $t(\hat{\beta}_j) = \dfrac{\hat{\beta}_j - \beta_j}{\hat{se}(\hat{\beta}_j)} \sim t[N-K] \quad \Rightarrow$ ***sample value*** $= t_0(\hat{\beta}_j) = \dfrac{\hat{\beta}_j - b_j}{\hat{se}(\hat{\beta}_j)}$.

- ***Decision Rules – left-tail t-test:***

Reject $\mathbf{H_0}$ if $t_0 < -t_\alpha[N-K]$ *or* ***left-tail p-value*** for $t_0 = \Pr(t < t_0) < \alpha$;

Retain $\mathbf{H_0}$ if $t_0 \ge -t_\alpha[N-K]$ *or* ***left-tail p-value*** for $t_0 = \Pr(t < t_0) \ge \alpha$.

- ***Decision Rules – right-tail t-test:***

Reject $\mathbf{H_0}$ if $t_0 > t_\alpha[N-K]$ *or* ***right-tail p-value*** for $t_0 = \Pr(t > t_0) < \alpha$;

Retain $\mathbf{H_0}$ if $t_0 \le t_\alpha[N-K]$ *or* ***right-tail p-value*** for $t_0 = \Pr(t > t_0) \ge \alpha$.

2. Tests of *One* Linear Restriction on *Two or More* Coefficients

H_0 specifies ***only one*** linear restriction on ***two or more*** regression coefficients.

❑ *Two-Tail* Tests of *One* Linear Restriction on *Two* Coefficients

- ***Example:*** H_0: $c_j\beta_j + c_h\beta_h = c_0$ versus H_1: $c_j\beta_j + c_h\beta_h \neq c_0$.
- ***Use:*** either a ***two-tail*** **t-test** or an **F-test**.
- ***Test Statistics:***

t-statistic

$$\Rightarrow\ t\left(c_j\hat{\beta}_j + c_h\hat{\beta}_h\right) = \frac{(c_j\hat{\beta}_j + c_h\hat{\beta}_h) - (c_j\beta_j + c_h\beta_h)}{\hat{se}(c_j\hat{\beta}_j + c_h\hat{\beta}_h)} \sim t[N-K] = t[N-K_1]$$

where: $\hat{se}(c_j\hat{\beta}_j + c_h\hat{\beta}_h) = \sqrt{\hat{Var}(c_j\hat{\beta}_j + c_h\hat{\beta}_h)}$

$$\hat{Var}(c_j\hat{\beta}_j + c_h\hat{\beta}_h) = c_j^2\hat{Var}(\hat{\beta}_j) + c_h^2\hat{Var}(\hat{\beta}_h) + 2c_jc_h\hat{Cov}(\hat{\beta}_j, \hat{\beta}_h).$$

sample value $= t_0\left(c_j\hat{\beta}_j + c_h\hat{\beta}_h\right) = \dfrac{(c_j\hat{\beta}_j + c_h\hat{\beta}_h) - c_0}{\hat{se}(c_j\hat{\beta}_j + c_h\hat{\beta}_h)}$.

❑ ***Two-Tail*** **Tests of** ***One*** **Linear Restriction on** ***Two*** **Coefficients (continued)**

F-statistic

$$\Rightarrow \quad F\left(c_j\hat{\beta}_j + c_h\hat{\beta}_h\right) = \frac{\left[(c_j\hat{\beta}_j + c_h\hat{\beta}_h) - (c_j\beta_j + c_h\beta_h)\right]^2}{\hat{Var}(c_j\hat{\beta}_j + c_h\hat{\beta}_h)} \sim F[1, N-K] = F[1, N-K_1]$$

where: $\hat{Var}(c_j\hat{\beta}_j + c_h\hat{\beta}_h) = c_j^2\hat{Var}(\hat{\beta}_j) + c_h^2\hat{Var}(\hat{\beta}_h) + 2c_jc_h\hat{Cov}(\hat{\beta}_j, \hat{\beta}_h)$.

sample value $= F_0\left(c_j\hat{\beta}_j + c_h\hat{\beta}_h\right) = \dfrac{\left[(c_j\hat{\beta}_j + c_h\hat{\beta}_h) - c_0\right]^2}{\hat{Var}(c_j\hat{\beta}_j + c_h\hat{\beta}_h)}$

- ***Decision Rules – Two-Tail Tests of One Coefficient Restriction on Two Coefficients:***

Reject **H₀** if $|t_0| > t_{\alpha/2}[N-K]$ *or* ***two-tail p-value*** for $t_0 = \Pr(|t| > |t_0|) < \alpha$;
$F_0 > F_\alpha[1, N-K]$ *or* ***p-value*** for $F_0 = \Pr(F > F_0) < \alpha$.

Retain **H₀** if $|t_0| \le t_{\alpha/2}[N-K]$ *or* ***two-tail p-value*** for $t_0 = \Pr(|t| > |t_0|) \ge \alpha$;
$F_0 \le F_\alpha[1, N-K]$ *or* ***p-value*** for $F_0 = \Pr(F > F_0) \ge \alpha$.

❑ ***One-Tail*** **Tests of *One* Linear Restriction on *Two* Coefficients**

- ***Examples:*** H_0: $c_j\beta_j + c_h\beta_h = c_0$ vs. H_1: $c_j\beta_j + c_h\beta_h < c_0$ a ***left-tail*** **test**
  H_0: $c_j\beta_j + c_h\beta_h = c_0$ vs. H_1: $c_j\beta_j + c_h\beta_h > c_0$ a ***right-tail*** **test**

- ***Use:*** a ***one-tail*** **t-test**.

- ***Test Statistic:***

$$\Rightarrow \; t\left(c_j\hat{\beta}_j + c_h\hat{\beta}_h\right) = \frac{(c_j\hat{\beta}_j + c_h\hat{\beta}_h) - (c_j\beta_j + c_h\beta_h)}{\hat{se}(c_j\hat{\beta}_j + c_h\hat{\beta}_h)} \sim t[N-K] = t[N-K_1]$$

where $\hat{se}(c_j\hat{\beta}_j + c_h\hat{\beta}_h) = \sqrt{\hat{Var}(c_j\hat{\beta}_j + c_h\hat{\beta}_h)}$

$$\hat{Var}(c_j\hat{\beta}_j + c_h\hat{\beta}_h) = c_j^2\hat{Var}(\hat{\beta}_j) + c_h^2\hat{Var}(\hat{\beta}_h) + 2c_jc_h\hat{Cov}(\hat{\beta}_j, \hat{\beta}_h).$$

sample value $= t_0\left(c_j\hat{\beta}_j + c_h\hat{\beta}_h\right) = \dfrac{(c_j\hat{\beta}_j + c_h\hat{\beta}_h) - c_0}{\hat{se}(c_j\hat{\beta}_j + c_h\hat{\beta}_h)}.$

❑ ***One-Tail* Tests of *One* Linear Restriction on *Two* Coefficients** **(continued)**

- ***Decision Rules – left-tail t-test:***

Reject* H₀** if $t_0 < -t_\alpha[N-K]$ *or* ***left-tail p-value for $t_0 = \Pr(t < t_0) < \alpha$;

Retain* H₀** if $t_0 \geq -t_\alpha[N-K]$ *or* ***left-tail p-value for $t_0 = \Pr(t < t_0) \geq \alpha$.

- ***Decision Rules – right-tail t-test:***

Reject* H₀** if $t_0 > t_\alpha[N-K]$ *or* ***right-tail p-value for $t_0 = \Pr(t > t_0) < \alpha$;

Retain* H₀** if $t_0 \leq t_\alpha[N-K]$ *or* ***right-tail p-value for $t_0 = \Pr(t > t_0) \geq \alpha$.

3. ***Tests*** **of** ***Two or More*** **Linear Coefficient Restrictions**

H_0 specifies ***two or more*** **linear coefficient restrictions**.

- ***Example:*** H_0: $\beta_2 = \beta_4$ *and* $\beta_3 = \beta_5$
  H_1: $\beta_2 \neq \beta_4$ *and/or* $\beta_3 \neq \beta_5$

- ***Use:*** a **general F-test**; only an F-test can be used to test jointly *two or more* coefficient restrictions.

- ***Test Statistics:*** Either of the following two forms of the general F-statistic.

$$\Rightarrow \quad F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N - K)}.$$

$$\Rightarrow \quad F = \frac{(R_U^2 - R_R^2)/(df_0 - df_1)}{(1 - R_U^2)/df_1} = \frac{(R_U^2 - R_R^2)/(K - K_0)}{(1 - R_U^2)/(N - K)}.$$

$df_0 = N - K_0$ = degrees-of-freedom for RSS_0; $df_1 = N - K$ = degrees-of-freedom for RSS_1.

number of restrictions specified by H_0 = $q = df_0 - df_1 = K - K_0$

- ***Null distribution:*** $F \sim F[df_0 - df_1, df_1] = F[K - K_0, N - K]$.

- ***Sample value*** of F-statistic under $H_0 = F_0$.

❑ ***Tests* of *Two or More* Linear Coefficient Restrictions (continued)**

- ***Decision Rules:***

Reject* H₀** if $F_0 > F_\alpha[df_0 - df_1, df_1] = F_\alpha[K - K_0, N - K]$ *or* ***p-value for $F_0 = \Pr(F > F_0) < \alpha$.

Retain* H₀** if $F_0 \le F_\alpha[df_0 - df_1, df_1] = F_\alpha[K - K_0, N - K]$ *or* ***p-value for $F_0 = \Pr(F > F_0) \ge \alpha$.

❑ ***Tests* of *Two* Linear Coefficient Restrictions: An Example**

- ***Example:*** H_0: $\beta_2 = \beta_4$ *and* $\beta_3 = \beta_5$
  H_1: $\beta_2 \neq \beta_4$ *and/or* $\beta_3 \neq \beta_5$

- ***Unrestricted model*** is given by PRE (1):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i \qquad K = 6 \qquad (1)$$

- ***Restricted model*** is given by PRE (2): set $\beta_4 = \beta_2$ *and* $\beta_5 = \beta_3$ in PRE (1):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i \qquad (1)$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_2 X_{4i} + \beta_3 X_{5i} + u_i$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 (X_{2i} + X_{4i}) + \beta_3 (X_{3i} + X_{5i}) + u_i \qquad K_0 = 4 \qquad (2)$$

- ***Number of restrictions*** that Restricted model (2) imposes on Unrestricted model (1):

  number of restrictions specified by $H_0 = q = K - K_0 = 6 - 4 = 2$.

❑ ***Tests* of *Two* Linear Coefficient Restrictions: An Example** (continued)

- OLS estimation of (1) yields the ***unrestricted SRE*** (1*):

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \hat{\beta}_5 X_{5i} + \hat{u}_i \qquad (1^*)$$

$$RSS_1 = \sum_{i=1}^{N} \hat{u}_i^2 = \hat{u}^T \hat{u} \quad \text{with} \quad df_1 = N - K = N - 6 .$$

- OLS estimation of (2) yields the ***restricted SRE*** (2*):

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 (X_{2i} + X_{4i}) + \tilde{\beta}_3 (X_{3i} + X_{5i}) + \tilde{u}_i \qquad (2^*)$$

$$\tilde{\beta}_4 = \tilde{\beta}_2 \quad \text{and} \quad \tilde{\beta}_5 = \tilde{\beta}_3$$

$$RSS_0 = \sum_{i=1}^{N} \tilde{u}_i^2 = \tilde{u}^T \tilde{u} \quad \text{with} \quad df_0 = N - K_0 = N - 4 .$$

- Substitute values of RSS_1, RSS_0, df_1 and df_0 into formula for general F-statistic:

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N - K)}$$

- Apply decision rule.

♦ Example of Tests of *Two* Linear Coefficient Restrictions: *Stata* Commands

- ***Example:*** H_0: $\beta_2 = \beta_4$ *and* $\beta_3 = \beta_5$
  H_1: $\beta_2 \neq \beta_4$ *and/or* $\beta_3 \neq \beta_5$

- ***Unrestricted model*** is given by PRE (1):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i \quad (1)$$

- Estimate ***unrestricted model*** given by PRE (1) by OLS using **Stata *regress*** command:

```
regress y x1 x2 x3 x4 x5
```

- Perform **joint F-test** of H_0 versus H_1 using **Stata *test*** command:

```
test x2 = x4
test x3 = x5, accumulate
```

or

```
test x2 - x4 = 0
test x3 - x5 = 0, accumulate
```

♦ **<u>Example of Tests of *One* Linear Coefficient Restriction: *Stata* Commands</u>**

- Perform **individual two-tail t-test** of H_0: $\beta_2 = \beta_4$ versus H_1: $\beta_2 \neq \beta_4$ using **Stata *lincom*** command:

```
lincom x2 - x4
```

or

```
lincom _b[x2] - _b[x4]
```

- Perform **individual F-test** of H_0: $\beta_2 = \beta_4$ versus H_1: $\beta_2 \neq \beta_4$ using **Stata *test*** command:

```
test x2 = x4
```

or

```
test x2 - x4 = 0
```

<u>Example of Tests of *One* Linear Coefficient Restriction: *Stata* Commands</u> (continued)

- How is $\mathbf{t_0}$ **statistic** produced by ***lincom*** command related to $\mathbf{F_0}$ **statistic** produced by ***test*** command?

lincom command computes:

$$t_0\left(\hat{\beta}_2 - \hat{\beta}_4\right) = \frac{(\hat{\beta}_2 - \hat{\beta}_4) - 0}{\hat{se}(\hat{\beta}_2 - \hat{\beta}_4)} = \frac{(\hat{\beta}_2 - \hat{\beta}_4)}{\hat{se}(\hat{\beta}_2 - \hat{\beta}_4)} \sim t[N-6] \text{ under } H_0$$

test command computes:

$$F_0\left(\hat{\beta}_2 - \hat{\beta}_4\right) = \frac{\left[(\hat{\beta}_2 - \hat{\beta}_4) - 0\right]^2}{\hat{Var}(\hat{\beta}_2 - \hat{\beta}_4)} = \frac{(\hat{\beta}_2 - \hat{\beta}_4)^2}{\hat{Var}(\hat{\beta}_2 - \hat{\beta}_4)} \sim F[1, N-6] \text{ under } H_0$$

- **Relationship between t and F tests**: they **yield *identical* inferences** because

$$F_0\left(\hat{\beta}_2 - \hat{\beta}_4\right) = \left(t_0(\hat{\beta}_2 - \hat{\beta}_4)\right)^2 \quad \text{and} \quad F_\alpha[1, N-6] = \left(t_{\alpha/2}[N-6]\right)^2$$

p-value for $\mathbf{F_0}$ = ***two-tail p-value*** for $\mathbf{t_0}$