8.2 Omitting some regressors

Most common couple model-submodel

$M: \quad Y|Z \sim (X\beta, \sigma^2 I_n), \quad X = (\underset{k_0}{X^0}, \underset{k_1}{X^1})$ columns

$M_0: \quad \sim (X^0\beta^0, \sigma^2 I_n) \qquad \beta = (\beta^{0T}, \beta^{1T})^T$

$$0 < \text{rank}(X^0) = k_0 < k = \text{rank}(X) < n$$

$$\text{rank}(X^1) = k_1 = k - k_0$$

(all involved matrices of full column rank)

$H_0: \quad E(Y|Z) \in \mathcal{M}(X^0) \qquad , \beta_1 = 0$

$H_1: \quad \in \underbrace{\mathcal{M}(X^0, X^1) \setminus \mathcal{M}(X^0)}_{\neq \mathcal{M}(X^1)\,!} \quad , \beta_1 \neq 0$

Remark: $E(Y|Z) \in \mathcal{M}(X^0, X^1) \setminus \mathcal{M}(X^0)$

= also some columns from X^1 are needed (on top of columns from X^0) to capture (by a linear combination) $E(Y|Z)$

Lemma 8.3 Effect of omitting some regressors

Consider a couple (model - submodel), where the submodel is obtained by omitting some regressors from the model. The following then holds.

(i) If $\mathcal{M}(X^1) \perp \mathcal{M}(X^0)$ then

$$D = X^1(X^{1T}X^1)^{-1}X^{1T}Y =: \hat{Y}^1,$$

which are the fitted values from a linear model $Y/Z \sim (X^1\beta^1, \sigma^2 I_n)$.

Proof: $M^0 := I - X^0(X^{0T}X^0)^{-1}X^{0T}$ (residual projection matrix for model M_0)

$$\mathcal{M}(\underbrace{X^0, X^1}_{X}) = \mathcal{M}(\underbrace{X^0, M^0X^1}_{\text{mutually orthogonal columns}})$$

since $M^0X^1 = X^1 - X^0(X^{0T}X^0)^{-1}X^{0T}X^1$

= combination of columns of X^0, X^1

$$\mathbb{P} = (\overbrace{\underset{k_0}{Q^0}, \underset{k_1}{Q^1}}^{k}, \underset{n-k}{N})$$

$\underbrace{Q^0}_{\mathcal{M}(X^0)}$ $\underbrace{Q^1}_{\mathcal{M}(Q^1) = \mathcal{M}(M^0X^1)}$ because X^0 and M^0X^1 have orthogonal columns

---

$D =$ projection of Y into $\mathcal{M}(Q^1) = \mathcal{M}(M^0X^1)$

- projection matrix to get D: "$H = X(X^TX)^{-1}X^T$", now "$X = M^0X^1$"

$$\to \quad D = \underbrace{M^0X^1(X^{1T}\overbrace{M^0M^0}^{M^0}X^1)^{-1}X^{1T}M^0}_{=: H^1 \text{ (project. matrix to } \mathcal{M}(Q^1))}\overbrace{Y}^{U^0} \quad (= \hat{Y} - \hat{Y}^0)$$

That is $D = \hat{Y} - \hat{Y}^0 = M^0X^1(X^{1T}\underbrace{M^0}_{M^0M^0}X^1)^{-1}X^{1T}\underbrace{U^0}_{M^0Y}$

If $\mathcal{M}(X^1) \perp \mathcal{M}(X^0) \Rightarrow M^0 X^1 = X^1$

$\uparrow$ project. matrix to $\mathcal{M}(X^0)^\perp$

Hence $\mathbb{D} = X^1 (X^{1T} X^1)^{-1} X^{1T} \mathbb{Y}$. $\boxed{\text{III}}$

(iii) If for given $\mathbb{Z}$, the conditional distribution $\mathbb{Y}/\mathbb{Z}$ is continuous, i.e., has a density with respect to the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}_n)$ then $\mathbb{D} \neq 0_n$ and $SS_e^0 - SS_e > 0$ almost surely.

Proof: $\mathbb{D} = \mathbb{0} \iff$ projection of $\mathbb{Y}$ into $\mathcal{M}(Q^1)$ (which vector dim $= \underbrace{k-k_0}_{k_1} > 0$) is a zero vector

$\iff \mathbb{Y} \in \underbrace{\mathcal{M}(Q^1)^\perp}_{\text{vector dimension} = n-k_1 < n}$

$\mathbb{Y} \in \mathbb{R}^n$ and with continuous distribution

$P(\mathbb{Y} \in \mathcal{M}(Q^1)^\perp / \mathbb{Z}) = 0$.

Hence, almost surely, $\mathbb{D} \neq \mathbb{0}$, $SS_e^0 - SS_e > 0$. $\boxed{\text{III}}$

Remark: If $\mathbb{Y}/\mathbb{Z}$ has a continuous distribution then removal of regressors from the model always (almost surely) increases SS_e.

Question is, whether this increase is "important" (statistically significant)

$\rightarrow$ F-test.

8.3 Linear constraints

15

$Y/Z \sim (X\beta, \sigma^2 I_n)$, $\operatorname{rank}(X_{n\times k}) = k$

$L_{m\times k} = \begin{pmatrix} l_1^T \\ \vdots \\ l_m^T \end{pmatrix}$ given matrix with linearly independent rows, $m < k$

$\theta^0 \in \mathbb{R}^m$ given vector

Can we write $E(Y/Z) = X\beta$ where β satisfies $L\beta = \theta^0$?

$\equiv$? $E(Y/Z) \in \mathcal{M}(X; L\beta = \theta^0)$

$= \{v : v = X\beta, L\beta = \theta^0\} \subset \mathbb{R}^n$

Def 8.2 Submodel given by linear constraints

16

We say that the model M_0 is a submodel given by linear constraints $L\beta = \theta^0$ of model $M: Y/Z \sim (X\beta, \sigma^2 I_n)$, $\operatorname{rank}(X_{n\times k}) = k$, if the response expectation $E(Y/Z)$ under the model M_0 is assumed to lie in a space $\mathcal{M}(X; L\beta = \theta^0)$, where $L_{m\times k}$ is a real matrix with m linearly independent rows, $m<k$ and $\theta^0 \in \mathbb{R}^m$ is a given vector.

Notation: A submodel given by linear constraints will be denoted as

$$M_0: Y/Z \sim (X\beta, \sigma^2 I_n), L\beta = \theta^0$$

Def 8.3 Fitted values, residuals, residual sum of squares, rank of the model and residual degrees of freedom in a submodel given by linear constraints

Let $b^0 \in \mathbb{R}^k$ minimizes $SS(\beta) = \|Y - X\beta\|^2$ over $\beta \in \mathbb{R}^k$ subject to $L\beta = \theta^0$. For the model $M_0: Y/Z \sim (X\beta, \sigma^2 I_n), L\beta = \theta^0$, the following quantities are defined as follows

Fitted values: $\hat{Y}^0 := Xb^0$

Residuals: $U^0 := Y - \hat{Y}^0$

Residual sum of squares: $SS_e^0 := \|U^0\|^2$

Rank of the model: $r_0 = k - m$

Residual degrees of freedom: $\nu_e^0 = n - r_0$

Note: The fitted values could also be defined as $\hat{Y}^0 = \underset{\tilde{Y} \in \mathcal{M}(X; L\beta = \theta^0)}{\operatorname{argmin}} \|Y - \tilde{Y}\|^2$.

18

Theorem 8.4 On a submodel given by linear constraints

Let $M_0: Y/Z \sim (X\beta, \sigma^2 I_n), L\beta = \theta^0$ be a submodel given by linear constraints of a model $M: Y/Z \sim (X\beta, \sigma^2 I_n)$

Then

(i) There is a unique minimizer b^0 to $SS(\beta) = \|Y - X\beta\|^2$ subject to $L\beta = \theta^0$ and is given as

$$b^0 = \hat{\beta} - (X^T X)^{-1} L^T \{L(X^T X)^{-1} L^T\}^{-1} (L\hat{\beta} - \theta^0),$$

where $\hat{\beta} = (X^T X)^{-1} X^T Y$ is the (unconstrained) LSE of β.

(ii) The fitted values $\hat{Y}^0$ can be expressed as

$$\hat{Y}^0 = \hat{Y} - X(X^T X)^{-1} L^T \{L(X^T X)^{-1} L^T\}^{-1} (L\hat{\beta} - \theta^0).$$

(iii) The vector $D = \hat{Y} - \hat{Y}^0$ satisfies

$$\|D\|^2 = SS_e^0 - SS_e = (L\hat{\beta} - \theta^0)^T \{L(X^T X)^{-1} L^T\}^{-1} (L\hat{\beta} - \theta^0).$$

Before proof:

- points (i) and (ii) just technical
- do <u>not</u> learn by heart
- point (iii) $\equiv$ "numerator" of the Wald F-statistic developed earlier (Theorem 6.2) to test $L\beta = \theta^0$

Proof: Remark – matrix $L(X^TX)^{-1}L^T$ is invertible (Theorem 2.5 Gauss-Markov for lin. comb.)

Try to look for $\hat{Y}^0 = X b^0$ such that b^0 minimizes $SS(\beta) = \|Y - X\beta\|^2$ subject to $L\beta = \theta^0$

Lagrange multipliers

cosmetic

$$\varphi(\beta, \lambda) = SS(\beta) + 2\lambda^T(L\beta - \theta^0)$$

$$\frac{\partial \varphi}{\partial \beta} = \underbrace{-2X^T(Y - X\beta)}_{\to \text{normal equations (see Chapter 2)}} + 2L^T\lambda$$

$$\frac{\partial \varphi}{\partial \lambda} = 2(L\beta - \theta^0)$$

$$\frac{\partial \varphi}{\partial \beta} = 0 \iff \underbrace{X^TX\beta = X^TY - L^T\lambda}_{\text{is always (for any } \lambda) \text{ solvable}}$$

$\mathcal{M}(L^T) \subset \mathbb{R}^k = \mathcal{M}(X^T) = \mathcal{M}(X^TX)$ RHS $\in \mathcal{M}(X^TX) = \mathcal{M}(X^T)$

LHS $\in \mathcal{M}(X^T) = \mathcal{M}(X^TX)$

For any $\lambda \in \mathbb{R}^m$ $\quad \frac{\partial \varphi}{\partial \beta} = 0$ for

$$\beta = b^0(\lambda) = (X^T X)^{-1}(X^T Y - L^T\lambda) =$$

$$= \underbrace{(X^T X)^{-1} X^T Y}_{\hat{\beta}} - (X^T X)^{-1} L^T \lambda = \hat{\beta} - (X^T X)^{-1} L^T \lambda$$

$$\frac{\partial \varphi}{\partial \lambda} = 0 \iff L\,b^0(\lambda) = \theta^0$$

$$L\hat{\beta} - L(X^T X)^{-1} L^T \lambda = \theta^0$$

$$\underbrace{L(X^T X)^{-1} L^T}_{\text{invertible}} \lambda = L\hat{\beta} - \theta^0$$

$$\lambda = \left(L(X^T X)^{-1} L^T\right)^{-1}(L\hat{\beta} - \theta^0)$$

$$\Rightarrow b^0 = \hat{\beta} - (X^T X)^{-1} L^T \underbrace{\left(L(X^T X)^{-1} L^T\right)^{-1}(L\hat{\beta} - \theta^0)}_{\lambda}$$

III (i)

$$\hat{Y}^0 = X b^0 = \underbrace{X\hat{\beta}}_{\hat{Y}} - X(X^T X)^{-1} L^T \left(L(X^T X)^{-1} L^T\right)^{-1}(L\hat{\beta} - \theta^0)$$

$$= \hat{Y} - X(X^T X)^{-1} L^T \left(L(X^T X)^{-1} L^T\right)^{-1}(L\hat{\beta} - \theta^0)$$

III (ii)

$$D = \hat{Y} - \hat{Y}^0 = X(X^T X)^{-1} L^T \left(L(X^T X)^{-1} L^T\right)^{-1}(L\hat{\beta} - \theta^0)$$

from previous page:

$$D = \hat{Y} - \hat{Y}^0 = X(X^TX)^{-1}L^T\overbrace{(L(X^TX)^{-1}L^T)^{-1}}^{V^{-1}}(L\hat{\beta}-\theta^0)$$

$$\|D\|^2 = (L\hat{\beta}-\theta^0)^T V^{-1} \underbrace{\boxed{L\underbrace{(X^TX)^{-1}X^TX}_{I}(X^TX)^{-1}L^T}}_{V} V^{-1}(L\hat{\beta}-\theta^0) =$$

$$= (L\hat{\beta}-\theta^0)^T V^{-1}(L\hat{\beta}-\theta^0) =$$

$$= (L\hat{\beta}-\theta^0)^T (L(X^TX)^{-1}L^T)^{-1}(L\hat{\beta}-\theta^0)$$

Finally

$$SSe^0 = \|Y-\hat{Y}^0\|^2 =$$

$$= \|\underbrace{Y-\hat{Y}}_{U\in\mathcal{M}(X)^\perp} + \underbrace{X(X^TX)^{-1}L^T(L(X^TX)^{-1}L^T)^{-1}(L\hat{\beta}-\theta^0)}_{D\in\mathcal{M}(X)}\|^2$$

$$= \|U\|^2 + \|D\|^2 = SSe + \|D\|^2.$$

Hence $\|D\|^2 = SSe^0 - SSe$. ▯ (iii)

Testing a linear constraint

$\equiv$ testing a submodel

8.3.1 F-statistic to verify a set of linear constraints

H_0: $\mathbb{E}(Y/Z) \in \mathcal{M}(X; L\beta = \theta^0)$

H_1: $\mathbb{E}(Y/Z) \in \mathcal{M}(X) \setminus \mathcal{M}(X; L\beta = \theta^0)$

$H_0: L\beta = \theta^0 \in \mathbb{R}^m$

$H_1: L\beta \neq \theta^0$

Form of the F-statistic on submodel by Theorem 8.1:

$$F_0 = \frac{\frac{SSe^0 - SSe}{(n-r_0)-(n-k)}}{\frac{SSe}{n-k}} = \frac{\frac{\|D\|^2}{k-r_0}}{\frac{SSe}{n-k}} =$$

$$= \frac{\frac{(L\hat{\beta}-\theta^0)^T \{L(X^TX)^{-1}L^T\}^{-1}(L\hat{\beta}-\theta^0)}{m}}{\frac{SSe}{n-k}} =$$

MSe

$$= \frac{1}{m}(L\hat{\beta}-\theta^0)^T\{MSe\,L(X^TX)^{-1}L^T\}^{-1}(L\hat{\beta}-\theta^0)$$

$\hat{\theta} := L\hat{\beta}$

$$\Rightarrow F_0 = \frac{1}{m}(\hat{\theta}-\theta^0)^T\{MSe\,L(X^TX)^{-1}L^T\}^{-1}(\hat{\theta}-\theta^0)$$

= Wald-type statistic Q_0 derived in Theorem 6.2 to test $H_0: \theta = \theta^0$
$\equiv L\beta = \theta^0$

Theorem 6.2: under normality (and H_0)

$$F_0 \sim F_{\underset{\substack{\uparrow \\ k-r_0}}{m}, n-k}$$

8.3.2 t-statistic to verify a linear constraint

$L = \ell^T$, $\ell \neq 0$, $\ell \in \mathbb{R}^k$, $\theta^0 \in \mathbb{R}$

H_0: $\mathbb{E}(Y|Z) \in \mathcal{M}(X; \ell^T\beta = \theta^0)$

H_1: $\mathbb{E}(Y|Z) \in \mathcal{M}(X) \setminus \mathcal{M}(X; \ell^T\beta = \theta^0)$

H_0: $\ell^T\beta = \theta^0 \in \mathbb{R}$

H_1: $\ell^T\beta \neq \theta^0$

The statistic F_0 becomes $\qquad \hat{\theta} := \ell^T\hat{\beta}$

$$F_0 = \frac{1}{1}(\hat{\theta}-\theta^0)^T \{ MSe\, \ell^T(X^TX)^{-1}\ell \}^{-1} (\hat{\theta}-\theta^0)$$

$$= \left(\frac{\hat{\theta}-\theta^0}{\sqrt{MSe\, \ell^T(X^TX)^{-1}\ell}} \right)^2 =: T_0^2$$

Theorem 6.2 : under normality (and H_0)

$$T_0^2 \sim \mathcal{F}_{1, n-k}$$

$$\Rightarrow T_0 \sim t_{n-k}$$

8.4 Overall F-test

Lemma 8.5 Overall F-test

Assume a normal linear model $Y|X \sim N(X\beta, \sigma^2 I_n)$, rank $(X_{n \times k}) = r > 1$ where $\mathbb{1} \in \mathcal{M}(X)$. Let R^2 be its coefficient of determination. The submodel F-statistic to compare model $M: Y|X \sim N(X\beta, \sigma^2 I_n)$ and the only intercept model $M_0: Y|X \sim N(\mathbb{1}_n \mu, \sigma^2 I_n)$ takes the form

$$F_0 = \frac{R^2}{1-R^2} \cdot \frac{n-r}{r-1}.$$

Proof (easy algebra...)

intercept only model

- $R^2 = 1 - \frac{SS_e}{SS_T}$, Lemma 7.1: $SS_T = SS_e^0$

- $R^2 = 1 - \frac{SS_e}{SS_e^0} = \frac{SS_e^0 - SS_e}{SS_e^0}$, $1 - R^2 = \frac{SS_e}{SS_e^0}$

- $F_0 \overset{\text{Theorem 8.1}}{=} \frac{\frac{SS_e^0 - SS_e}{r-1}}{\frac{SS_e}{n-r}} = \frac{n-r}{r-1} \cdot \frac{SS_e^0 - SS_e}{SS_e} = \frac{n-r}{r-1} \cdot \frac{R^2}{1-R^2}$.

$\boxtimes$

→ F-statistic on bottom of R output

→ sometimes called as "goodness-of-fit" test which is not correct

→ test says nothing on a quality of the model w.r.t. $E(Y|Z)$

→ If significant ≡ at least one included regressor is significantly associated with $E(Y|Z)$.