IX. Checking Model Assumptions

Data (Y_i, Z_i), $i = 1, \dots, n$, $Z_i \in \mathcal{Z} \subseteq \mathbb{R}^p$

Possibly two sets of regressors

$$X_i = t_X(Z_i) \quad , \quad t_X : \mathbb{R}^p \to \mathbb{R}^k$$

$$V_i = t_V(Z_i) \quad , \quad t_V : \mathbb{R}^p \to \mathbb{R}^l$$

→ two model matrices

$$X = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix} = (X^0, X^1, \dots, X^{k-1}) \text{, mostly } X^0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$V = \begin{pmatrix} V_1^T \\ \vdots \\ V_n^T \end{pmatrix} = (V^1, \dots, V^l)$$

Assumed linear model for $Y = (Y_1, \dots, Y_n)^T$

$E(Y|Z) = X\beta$ for some β

$\operatorname{var}(Y|Z) = \sigma^2 I_n$ for some σ^2

sometimes also assumed that $Y|Z \sim N$

Assumptions behind $Y|Z \sim N(X\beta, \sigma^2 I_n)$

$\varepsilon := Y - X\beta = (Y_1 - X_1^T\beta, \dots, Y_n - X_n^T\beta)^T = (\varepsilon_1, \dots, \varepsilon_n)^T$

remember: $X_i = t_X(Z_i)$

1. Correct regression function

$E(Y_i|Z_i) = X_i^T\beta$ for some β

$\equiv E(\varepsilon_i|Z_i) = 0 \quad \forall i$

↑ no dependence on Z_i

2. (Conditional) homoscedasticity of errors

$\mathrm{var}(Y_i|Z_i) = \sigma^2$ for some σ^2

$\equiv \mathrm{var}(\varepsilon_i|Z_i) = \sigma^2 \quad \forall i$

↑ no dependence on Z_i

3. (Conditionally) uncorrelated / independent errors

$\mathrm{var}(Y|Z)$ is diagonal

4. (Conditionally) normal errors

$Y_i|Z_i \sim N$, $\varepsilon_i|Z_i \sim N \quad \forall i$

---

1. $E(\varepsilon|Z) = 0$
2. & 3. $\mathrm{var}(\varepsilon|Z) = \sigma^2 I_n$
4. $\varepsilon|Z \sim N$

---

also marginally, but now not that important

$1. \Rightarrow E\varepsilon_i = 0 \quad \forall i$

$2. \Rightarrow \mathrm{var}\,\varepsilon_i = \sigma^2 \quad \forall i$

$3. \Rightarrow \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$

$4. \Rightarrow \varepsilon_i \overset{iid}{\sim} N$

9.1 Model with added regressors

TECHNICAL SECTION

Let us assume two (competing) models:

$M: Y|Z \sim (X\beta, \sigma^2 I_n)$, $X_{n\times k}$

$M_g: Y|Z \sim (\underbrace{X\beta + V\gamma}_{G\cdot\begin{pmatrix}\beta\\ \gamma\end{pmatrix}}, \sigma^2 I_n)$, $V_{n\times l}$, $G = (X, V)$

Quantities derived while assuming model $M: Y|Z \sim (X\beta, \sigma^2 I_n)$

$b = (X^TX)^- X^T Y$ = any solution to normal equations $X^TX\beta = X^TY$

$\hat\beta = (X^TX)^{-1} X^TY$ = LSE of β if $\operatorname{rank}(X) = k$

$H = X(X^TX)^- X^T = (h_{it})_{i,t=1,\dots,n}$: hat matrix = projection to $\mathcal{M}(X)$

$\hat Y = HY = (\hat y_1, \dots, \hat y_n)^T$ = fitted values

$M = I_n - H = (m_{it})_{i,t=1,\dots,n}$ = projection to $\mathcal{M}(X)^\perp$

$U = Y - \hat Y = MY = (u_1, \dots, u_n)^T$ = residuals

$SSe = \|U\|^2$

Quantities derived while assuming model

$M_g: Y|Z \sim (X\beta + V_g, \sigma^2 I_n)$, $G = (\underset{k}{X}, \underset{l}{V})$

$(b_g^T, c_g^T)^T = (G^TG)^- G^T Y$ = any solution to normal eq. $G^TG \begin{pmatrix} \beta \\ g \end{pmatrix} = G^TY$

$(\hat{\beta}_g^T \; \hat{g}_g^T)^T = (G^TG)^{-1} G^TY$ = LSE of $\begin{pmatrix} \beta \\ g \end{pmatrix}$ if $\operatorname{rank}(G) = k+l$

$H_g = G(G^TG)^- G^T = (h_{g;it})_{i,t=1,\dots,n}$

$\hat{Y}_g = H_g Y = (\hat{y}_{g,1}, \dots, \hat{y}_{g,n})^T$

$M_g = I_n - H_g = (m_{g;it})_{i,t=1,\dots,n}$

$U_g = Y - \hat{Y}_g = M_g Y = (u_{g,1}, \dots, u_{g,n})^T$

$SS_{g,e} = \|U_g\|^2$

Lemma 9.1 Model with added regressor

Quantities derived while assuming model $M: Y|Z \sim (X\beta, \sigma^2 I_n)$ and quantities derived while assuming model $M_g: Y|Z \sim (X\beta + V\gamma, \sigma^2 I_n)$ are mutually in the following relationship.

$$\hat{Y}_g = \hat{Y} + MV(V^TMV)^-V^TU$$
$$= X b_g + V c_g \quad \text{for some } b_g \in \mathbb{R}^k, c_g \in \mathbb{R}^l.$$

Vectors b_g and c_g such that $\hat{Y}_g = Xb_g + Vc_g$ satisfy

$$c_g = (V^TMV)^-V^TU$$
$$b_g = b - (X^TX)^-X^TVc_g$$
for some $b = (X^TX)^-X^TY$.

Finally $SSe - SSe_g = \|MVc_g\|^2$.

Proof: • $\hat{Y}_g$ is projection of Y into $\mathcal{M}(X,V) = \mathcal{M}(X, \underbrace{MV}_{\perp})$

$$M = I_n - X(X^TX)^-X^T$$
$$MV = V - X(X^TX)^-X^TV$$

$\rightarrow \mathcal{M}(X, MV)$ is still a vector space being generated by columns from X, V.

• use "$H = X(X^TX)^-X^T$" to calculate H_g:

$$H_g = (X, MV)\begin{pmatrix} X^TX & \underbrace{X^TMV}_{0} \\ \underbrace{V^TMX}_{0} & V^TMV \end{pmatrix}^- \begin{pmatrix} X^T \\ V^TM \end{pmatrix} =$$

M·M

$$= (X, MV)\begin{pmatrix} (X^TX)^- & 0 \\ 0 & (V^TMV)^- \end{pmatrix}\begin{pmatrix} X^T \\ V^TM \end{pmatrix} = \underbrace{X(X^TX)^-X^T}_{H} + MV(V^TMV)^-V^TM$$

So that $\hat{\vec{y}}_g = H_g \vec{y} = \underbrace{H\vec{y}}_{\hat{\vec{y}}} + MV(V^TMV)^-V^T\underbrace{M\vec{y}}_{U} =$

$= \hat{\vec{y}} + MV(V^TMV)^-V^TU$ (*)

- The fitted values $\hat{\vec{y}}_g$ must lie in the corresponding regression space $\mathcal{M}(X, V)$
  → it must exist $b_g \in \mathbb{R}^k$, $c_g \in \mathbb{R}^l$ such that
  $\hat{\vec{y}}_g = Xb_g + Vc_g$

At the same time, $(b_g^T, c_g^T)^T$ must minimize the sum of squares of model M_g.

Side note: Proof of Lemma 2.1: vector $(b_g^T, c_g^T)^T$ minimizes the sum of squares $\Leftrightarrow$ $\Leftrightarrow$ it solves corresponding normal eqs.

Try to rewrite (*) to see what b_g and c_g could be.

↓ same arguments: $\hat{\vec{y}} = Xb \Leftrightarrow b = (X^TX)^-X^T\vec{y}$
(b solves normal eqs.)

$\Rightarrow$ (*) $\hat{\vec{y}}_g = \underbrace{Xb}_{\hat{\vec{y}}} + \underbrace{(I_n - X(X^TX)^-X^T)}_{M} V(V^TMV)^-V^TU =$

$=$ we need $X \cdot \square + V \cdot \square$ $X\underbrace{(b - (X^TX)^-X^TV(V^TMV)^-V^TU)}_{b_g} + V\underbrace{(V^TMV)^-V^TU}_{c_g}$

Finally $SSe - SSe_g = \|\hat{\vec{y}}_g - \hat{\vec{y}}\|^2 =$

$= \|MV\underbrace{(V^TMV)^-V^TU}_{c_g}\|^2 = \|MVc_g\|^2.$