# BIC in different models

1) Linear Model

$$Y_i = X_{i,A_0}' \beta_{A_0} + u_i$$

$$SSR_n(A) = \| Y_i - X_{i,A}' \hat{\beta}_{n,A}(A) \|^2$$

Define

$$BIC_n(A) = SSR_n(A) + |A| \log n$$ ( It's a measure of fit + penalty )

$$\hat{A}_n^{BIC} = \arg\min_A BIC_n(A)$$

Proof (of oracle property $P(\hat{A}_n^{BIC} = A_0) \to 1$ as $n \to \infty$. )

We need to show that

$$P(BIC_n(A) > BIC_n(A_0)) \to 1 \quad \text{as } \begin{matrix} n \to \infty \\ A \neq A_0 \end{matrix}.$$

- $$\begin{aligned} SSR_n(A_0) &= \frac{1}{n}\sum_{i=1}^n (Y_i - X_{i,A_0}' \hat{\beta}_{n,A_0}(A_0))^2 \\ &= \frac{1}{n}\sum_{i=1}^n (u_i - X_{i,A_0}'(\hat{\beta}_{n,A_0}(A_0) - \beta_{A_0}))^2 \\ &= \frac{1}{n}\sum_{i=1}^n u_i^2 - 2\left(\frac{1}{n}\sum_{i=1}^n X_{i,A_0} u_i\right)(\hat{\beta}_{n,A_0}(A_0) - \beta_{A_0}) \\ &\quad + (\hat{\beta}_{n,A_0}(A_0) - \beta_{A_0})'\left(\frac{1}{n}\sum_{i=1}^n X_{i,A_0} X_{i,A_0}'\right)(\hat{\beta}_{n,A_0}(A_0) - \beta_{A_0}) \\ &= Eu_i^2 - 2\,O(1)\,o_p(1) + o_p(1)\,O(1)\,o_p(1) \\ &= Eu_i^2 + o_p(1) \end{aligned}$$

 require $Ex_ix_i'$, $Eu_i^2 x_ix_i' < \infty$ and p.d. $Eu_i^2 < \infty$.

- $\frac{1}{n} SSR_n(A)$ s.t. $(A \cap A_0) \neq A_0$ ( omit relevant regressors )

$$\begin{aligned} \frac{1}{n} SSR_n(A) &= \frac{1}{n}\sum_{i=1}^n (Y_i - X_i' \hat{\beta}_n(A))^2 \\ &= \frac{1}{n}\sum_{i=1}^n u_i^2 + (\hat{\beta}_n(A) - \beta)'\left(\frac{1}{n}\sum_{i=1}^n x_ix_i'\right)(\hat{\beta}_n(A) - \beta) \\ &\quad - 2\left(\frac{1}{n}\sum_{i=1}^n x_iu_i\right)(\hat{\beta}_n(A) - \beta) \\ &= Eu_i^2 + \delta' E x_ix_i' \delta + o_p(1), \end{aligned}$$

where $\beta_{A_0} = \begin{pmatrix} \beta \\ 0 \end{pmatrix}$ ($\beta$: kept relevant; $0$: omitted relevant)

where $\hat{\beta}_n(A) - \beta \xrightarrow{p} \delta \neq 0$.

- $SSR_n(A)$ s.t. $A_0 \subset A$ (contain irrelevant regressors)

$$= \sum_{i=1}^{n} u_i^2 - 2\left(\sum_{i=1}^{n} X_{i,A_0} u_i\right)(\hat{\beta}_{n,A_0}(A_0) - \beta_{A_0}) + (\hat{\beta}_{n,A_0}(A_0) - \beta_{A_0})'\left(\frac{1}{n}\sum_{i=1}^{n} X_{i,A_0} X_{i,A_0}'\right)(\hat{\beta}_{n,A_0}(A_0) - \beta_{A_0})$$

$$= \frac{1}{n}\sum_{i=1}^{n} u_i^2 - 2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{i,A} u_i\right)\sqrt{n}(\hat{\beta}_{n,A} - \beta_A) + \sqrt{n}(\hat{\beta}_{n,A} - \beta_A)'\left(\frac{1}{n}\sum_{i=1}^{n} x_{i,A} x_{i,A}'\right)\sqrt{n}(\hat{\beta}_{n,A} - \beta_A)'$$

$$= \frac{1}{n}\sum_{i=1}^{n} u_i^2 + O_p(1) + o_p(1)$$

Then

- $P(BIC_n(A) > BIC_n(A_0))$ s.t. $A \cap A_0 = A_0$

$$= P\left(Eu_i^2 + \delta' E X_i X_i' \delta + o_p(1) + |A| \frac{\log n}{n} > Eu_i^2 + |A_0| \frac{\log n}{n} + o_p(1)\right)$$

$$= P\left((|A| - |A_0|)\frac{\log n}{n} + o_p(1) + \delta' E X_i X_i' \delta > 0\right)$$

$$\to 1.$$

- $P(BIC_n(A) > BIC_n(A_0))$ s.t. $A_0 \subset A$

$$= P\left(O_p(1) + |A| \frac{\log n}{n} > O_p(1) + |A_0| \frac{\log n}{n} + o_p(1)\right)$$

$$= P\left(O_p(1) + o_p(1) > (|A_0| - |A|) \log n\right)$$

$$\to 1.$$

2) M-estimator $\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \; Q_n(\theta)$ where $\hat{\theta}_n \xrightarrow{p} \theta_0$

(could be a pseudo-true parameter)

Consider 2nd order Taylor expansion around $\theta_0$

(A) $$Q_n(\hat{\theta}_n) = Q_n(\theta_0) + \frac{\partial Q_n(\theta_0)}{\partial \theta'}(\hat{\theta}_n - \theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)' \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\,\partial\theta'}(\hat{\theta}_n - \theta_0) + \underbrace{o(\|\hat{\theta}_n - \theta_0\|^2)}_{}\,,$$

$= o(1)\sqrt{n}(\hat{\theta}_n - \theta_0)'\sqrt{n}(\hat{\theta}_n - \theta_0)$

$= o(1) \quad O_p(1) \quad O_p(1)$

$= o_p(1)$

(B) $$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\underbrace{\Bigg[\underbrace{\frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'}}_{B_0}\Bigg]^{-1}\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta}}_{O_p(1)\text{ provided } B_0 \text{ is non-singular.}} + o_p(1)\,,$$

Hence

(C) $$\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta'}\sqrt{n}(\hat{\theta}_n - \theta_0) = -\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta'} B_0^{-1} \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta} + o_p(1)$$

(D) $$\sqrt{n}(\hat{\theta}_n - \theta_0)\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\,\partial\theta'}\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta'} B_0^{-1} \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta} + o_p(1)$$

Therefore

$$-2n\big(Q_n(\hat{\theta}_n) - Q_n(\theta_0)\big) = \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta'} B_0^{-1} \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta} + o_p(1)$$

$$\xrightarrow{d} \chi^2_m$$ $\rightarrow$ the degrees of freedom could change depending on the form of $\frac{\partial Q_n(\theta_0)}{\partial\theta}$

$$:= O_p(1)$$

now let $$A_n^{BIC} = \underset{A}{\operatorname{argmin}} \left\{ \underbrace{Q_{n,A}(\hat{\theta}_{n,A})}_{\text{measure of fit in these models.}} + |A| \frac{\log n}{n} \right\}$$

Then we wanna check $P(A_n^{BIC} = A_0) \to 1$ as $n \to \infty$.

- $A \cap A_0 \neq A_0$ ( include irrelevant regressors )

$$P\left( Q_{n,A}(\hat{\theta}_{n,A}) + |A| \frac{\log n}{n} > Q_{n,A_0}(\hat{\theta}_{n,A_0}) + |A_0| \frac{\log n}{n} \right)$$

$$= P\Big( \underbrace{Q_{n,A}(\hat{\theta}_{n,A}) - Q_{n,A_0}(\hat{\theta}_{n,A_0})}_{\text{misspecified} \Rightarrow \text{ converges to } \delta > 0 \text{ in prob.}} > (|A_0| - |A|) \underbrace{\frac{\log n}{n}}_{o(1)} \Big)$$

$$= P\big( \delta + o_p(1) > o(1) \big)$$

$$\to 1 \quad \text{as} \quad n \to \infty$$

- $A_0 \subset A$ (exclude relevant regressors )

$$P\left( 2n \left( Q_{n,A}(\hat{\theta}_{n,A}) - Q_{n,A_0}(\hat{\theta}_{n,A_0}) \pm Q_{n,A_0}(\theta_{A_0}) \right) > 2\cancel{n}(|A_0| - |A|) \frac{\log n}{\cancel{n}} \right)$$

$$= P\Big( \underbrace{2n\left( Q_{n,A}(\hat{\theta}_{n,A}) - Q_{n,A_0}(\theta_{A_0}) \right)}_{O_p(1)} + \underbrace{2n\left( Q_{n,A_0}(\theta_{A_0}) - Q_{n,A_0}(\hat{\theta}_{n,A_0}) \right)}_{O_p(1)} > 2(|A_0| - |A|) \log n \Big)$$

$$= P\big( O_p(1) > 2(|A_0| - |A|) \log n \big)$$

$$\to 1 \quad \text{as} \quad n \to \infty.$$