1. for a specific i.

$$-var(\hat{\varepsilon_i}) = -var[Y_i - \hat{Y_i}] = -E[(Y_i - \hat{Y_i})^2] + (E[Y_i - \hat{Y_i}])^2$$

$$var(\delta_i) = var[\hat{Y_i} - E[Y_i]] = E[(\hat{Y_i} - E[Y_i])^2] - (E[\hat{Y_i} - E[Y_i]])^2$$

Note that $(E[Y_i - \hat{Y_i}])^2 = (E[Y_i] - E[\hat{Y_i}])^2$

$$(E[\hat{Y_i} - E[Y_i]])^2 = (E[\hat{Y_i}] - E[Y_i])^2$$

we have, $-var(\hat{\varepsilon_i}) + var(\delta_i) = -E[(Y_i - \hat{Y_i})^2] + E[(\hat{Y_i} - E[Y_i])^2]$

$$= -E[RSS(\hat{Y_i})] + E[(\hat{Y_i} - E[Y_i])^2]$$

sum by i, we have.

$$E[RSS(\hat{Y})] - \sum_i var(\hat{\varepsilon_i}) + \sum_i var(\delta_i) = \sum_i E[(\hat{Y_i} - E[Y_i])^2]$$

$$= E\left[\sum_i (\hat{Y_i} - E[Y_i])^2\right]$$

$$= \sigma^2 p$$

2.

Denote $\hat{\varepsilon} \triangleq Y - \hat{Y} = (I-S)Y$, $\delta \triangleq \hat{Y} - E[Y] \triangleq \hat{Y} - \mu = SY - \mu$

$E[\hat{\varepsilon}] \triangleq \hat{\varepsilon}_\mu$. $E[\delta] \triangleq \delta_\mu$. $E[Y] = \mu$

Using results from 1. we only have to prove that.

$$\sigma^2(2tr(S) - n) = -\sum_{i=1}^n var(\hat{\varepsilon_i}) + \sum_{i=1}^n var(\delta_i)$$

$$= -E[(\hat{\varepsilon} - \hat{\varepsilon}_\mu)^T(\hat{\varepsilon} - \hat{\varepsilon}_\mu)] + E[(\delta - \delta_\mu)^T(\delta - \delta_\mu)]$$

We have $\hat{\varepsilon}_\mu = (I-S)E[Y] = (I-S)\mu$. $\hat{\varepsilon}-\hat{\varepsilon}_\mu = (I-S)(Y-\mu)$

$\delta_\mu = S\cdot E[Y]-\mu = (S-I)\mu$. $\delta-\hat{\delta}_\mu = S(Y-\mu)$

$$E[(\hat{\varepsilon}-\hat{\varepsilon}_\mu)^T(\hat{\varepsilon}-\hat{\varepsilon}_\mu)] = E[(Y-\mu)^T(I-S)^T(I-S)(Y-\mu)]$$

$$= E[(Y-\mu)^T(I-S-S^T+S^TS)(Y-\mu)] \quad --- (1)$$

$$E[(\delta-\delta_\mu)^T(\delta-\delta_\mu)] = E[(Y-\mu)^T S^T S (Y-\mu)] \quad --- (2)$$

$$-(1)+(2) = E[(Y-\mu)^T(-I+S+S^T)(Y-\mu)] \quad --- (3)$$

Use the trace trick.

$$(3) = \mathrm{tr}[(-I+S+S^T)\,\mathrm{Cov}(Y-\mu)] + (E[Y-\mu])^T(-I+S+S^T)E[Y-\mu]$$

$E[Y-\mu]=E[\varepsilon]=0$, $\mathrm{Cov}(Y-\mu) = E[(Y-\mu)(Y-\mu)^T] = E[\varepsilon\varepsilon^T] = \sigma^2\cdot I$

$$= \sigma^2\,\mathrm{tr}[-I+S+S^T] = \sigma^2\cdot(-n+2\,\mathrm{tr}[S]).$$

proved.

3. $S = X(X^TX)^{-1}X^T$.

$tr[S] = tr[X^TX(X^TX)^{-1}] = tr[I_p] = p$

$\therefore C_p = \frac{1}{\sigma^2} RSS(\hat{Y}) \quad p - n$

Note that $AIC = 2p - 2l$. if we treat σ^2 as a known constant, for normal linear model. $l = -\frac{1}{2}\{n \log \sigma^2 + \frac{1}{\sigma^2} RSS(\hat{Y}) + C\}$

$\therefore AIC = \frac{1}{\sigma^2} RSS(\hat{Y}) + 2p - n \log \sigma^2 + C$

(C is the constant for normalization)

Therefore C_p and AIC are quite similar for OLS model. despite the constant and coefficient for n.

4. Here we use $\hat{\sigma}_{ML} = \frac{RSS(\hat{\beta}_{ML})}{n}$ as the approximation of σ when computing AIC. Applying F-statistic. we have

$AIC(\hat{\beta}_{q+1}) - AIC(\hat{\beta}_q) = 2 - 2(l_{q+1} - l_q)$

where $l_q = -\frac{1}{2}\{n \log(RSS(\hat{\beta}_q)) + n - n \log n\}$

$\therefore P(AIC(\hat{\beta}_{q+1}) - AIC(\hat{\beta}_q) < 0)$

$= P(2 + n \log \frac{RSS(\hat{\beta}_{q+1})}{RSS(\hat{\beta}_q)} < 0)$

since $RSS(\hat{\beta_q}) - RSS(\hat{\beta}_{q+1}) \sim \sigma^2 \chi_1^2 \approx \frac{1}{n} RSS(\hat{\beta_q}) \cdot \chi_1^2$

therefore $\frac{RSS(\hat{\beta}_{q+1})}{RSS(\hat{\beta_q})} = 1 - \chi_1^2 / n$

$\therefore P\left(2 + n\log\frac{RSS(\hat{\beta}_{q+1})}{RSS(\hat{\beta_q})} < 0\right) = P\left(n\log\left(1 - \frac{\chi_1^2}{n}\right) < -2\right)$

$\overset{n\to\infty}{\approx} P\left(n \cdot \left(-\frac{\chi_1^2}{n}\right) < -2\right) = P(\chi_1^2 > 2) > 0.$

proved.

5. BIC : $-2\,l(\hat{\theta}) + p\log n$

$\therefore P\left(BIC(\hat{\beta}_{q+1}) - BIC(\hat{\beta_q}) < 0\right) = P\left(\log n + n\log\frac{RSS(\hat{\beta}_{q+1})}{RSS(\hat{\beta_q})} < 0\right)$

$\overset{n\to\infty}{\approx} P(\log n - \chi_1^2 < 0) = P(\chi_1^2 > \log n = \infty) = 0.$

proved.