Theorem: (Gauss-Markov)

Let $\hat{\vartheta} = (\Phi^T \Phi)^{-1} \Phi^T Y$. Then the OLS estimator $\hat{\vartheta}$ is

(1) An unbiased estimator of ϑ

(2) Optimal among all linear unbiased estimators (BLUE i.e. it has the smallest variance)

(3) The quantity:

$$\| \Phi(\Phi^T \Phi)^{-1} \Phi^T y - y \|^2$$

is an unbiased estimator of $(n-p)\sigma^2$.

In the Gaussian case (if the errors are normally distributed), these estimators (OLS) are also optimal among all unbiased estimators (not just linear!)

Proof: Let Vy be any linear estimator

$$\mathbb{E}_{\vartheta, \sigma^2}(Vy) = \mathbb{E}\Big(V(\Phi\vartheta + z)\Big) = V\Phi\vartheta + V \cancel{\mathbb{E}(z)}^{0}$$

So: $\mathbb{E}[Vy] = V\Phi\vartheta$. Vy is unbiased estimator iff $\vartheta = V\Phi\vartheta \ \forall \vartheta \Rightarrow$

$$\boxed{V\Phi = 1_p}$$ constrained for unbiasedness

OLS is a specific choise of V. We take

$V = (\Phi^T \Phi)^{-1} \Phi^T$

So $V\Phi = (\Phi^T\Phi)^{-1}\Phi^T\Phi = \mathbb{1}_p \Rightarrow$ OLS unbiased

Now we need to show that OLS has minimum variance. We calculate the MSE

$E_{\vartheta,\sigma^2}(\|Vy-\vartheta\|^2)$ (model $y = \Phi\vartheta + z$)

$E_{\vartheta,\sigma^2}(\|Vy-\vartheta\|^2) = E(\|V(\Phi\vartheta+z)-\vartheta\|^2)$

$= E(\|V\Phi\vartheta + Vz - \vartheta\|^2)$

$= E(\|(V\Phi - I_p)\vartheta + Vz\|^2)$

$= E(\|Vz\|^2) = \sigma^2 tr(VV^T) = \sigma^2 \sum_i (\sum_j V_{ij})^2$
$= \sigma^2 \|V\|^2$

Now we want to minimize

$$E(\|Vy-\vartheta\|^2) = \sigma^2\|V\|^2$$

subject to $V\Phi = I_p$

we use Lagrange multipliers

$L(V,\lambda) = \sigma^2\|V\|^2 - tr(\Lambda^T(V\Phi - I_p))$

Differentiating wrt V $\partial V = 0$

$\partial_V = 0$
$$\Rightarrow \frac{\partial}{\partial V}\left(\sigma^2 tr(VV^T - tr(\Lambda^T V \Phi))\right) = 0$$

$$\Rightarrow \quad 2\sigma^2 V = \Lambda \Phi^T$$

$$\Rightarrow \quad V = \frac{1}{2\sigma^2} \Lambda \Phi^T \quad (*)$$

we plug $(*)$ into the constraint

$$V \Phi = \mathbb{1}_p$$

$$\Rightarrow \quad \frac{1}{2\sigma^2} \Lambda \Phi^T \Phi = \mathbb{1}_p \Rightarrow \Lambda = 2\sigma^2 (\Phi^T \Phi)^{-1}$$

plugging back in we have

$$V = (\Phi^T \Phi)^{-1} \Phi^T$$

$\Rightarrow$ The only linear unbiased estimator V that minimizes the risk is the OLS estimator!

Distribution of the residual Sum of squares

Consider now the estimator

$$\| \Pi_{\Phi} Y - Y \|^2$$

orthogonal

the projection matrix onto the column space of Φ

orthogonal projection

$$\| \Pi_{\Phi} Y - Y \|^2 = \| \Pi_{\Phi} (\Phi \theta + \sigma Z) - (\Phi \theta + \sigma Z) \|^2$$

$$= \| (\Pi_{\Phi} \Phi - \Phi) \theta + \sigma (\Pi_{\Phi} Z - Z) \|^2$$

$$= \sigma^2 \| \Pi_{\Phi} Z - Z \|^2 \quad \text{if } Z \sim \mathcal{N}(0,1)$$

$$= \sigma^2 \| (\Pi_{\Phi} - \mathbb{1}_p) Z \|^2 \quad \Rightarrow \text{Cochran's theorem}$$

$$\mathbb{E} [\| (\Pi_{\Phi} - \mathbb{1}_p) Z \|^2 = \sum_{j=k+1}^{n} \mathbb{E} (Z_j^2) = n - k$$

if $Z \sim \mathcal{N}(0, I)$

$\Pi_{\Phi} Z \sim \mathcal{N}(0, \Pi_{\Phi})$

The squared norm $\| \Pi_{\Phi} Z \|^2$ follows a χ^2 with degrees of freedom equal to the rank of Π_Φ which is $n-p$

$$\frac{1}{\sigma^2} \| \Pi_\Phi Y - Y \|^2 \sim \chi^2_{n-p}$$

<u>Lemma</u>: Let $A \in O(n)$, z r.v in $\mathbb{R}^n$ with $\mathbb{E}(z)=0$ $\operatorname{cov}(z_i, z_j) = \Sigma_{ij}$

Then for $w = Az$ we have

$$\operatorname{cov}(w_i, w_j) = (A \Sigma A^T)_{ij}$$

<u>Proof</u>:

$$\operatorname{cov}(w_i, w_j) = \mathbb{E}(w_i w_j) = \mathbb{E}\left(\left(\sum_l A_{il} z_l \sum_k A_{jk} z_k\right) \right)$$

$$= \sum_{lk} A_{il} A_{jk} \mathbb{E}[z_l, z_k)$$

$$= \sum_{lk} A_{il} A_{jk} \Sigma_{lk}$$

$$= (A \Sigma A^T)_{ij}$$

$\|\Pi_\Phi z - z\|^2 = \|A^T \Pi_{1..p} A z - A^T A z\|^2$

$= \|A^T \Pi_{1p} A z - A z\|^2 = \sum_{j=p+1}^{n} (A z)_j^2$

$\Rightarrow$ Assume now $z \overset{iid}{\sim} \mathcal{N}(0,1)$

$S = (\Pi_\Phi Y, \|Y\|^2)$ is sufficient and complete

$(\Phi^T \Phi)^{-1} \Phi y$, $\|\Pi_\Phi Y - y\|^2$ are functions of S (i.e $\mathbb{E}(\cdot | S) = \cdot$)

$(\Phi^T \Phi)^{-1} \Phi^T y = (\Phi^T \Phi)^{-1} (\Pi_\Phi \Phi)^T y$

$= (\Phi^T \Phi)^{-1} \Phi^T \Pi_\Phi^T Y$

$= (\Phi^T \Phi)^{-1} \Phi^T (\Pi_\Phi Y)$

$\Rightarrow \|y\|^2 = \|\Pi_\Phi Y\|^2 + \|Y - \Pi_\Phi Y\|^2$

$\Rightarrow \|Y - \Pi_\Phi Y\|^2 = \|Y\|^2 - \|\Pi_\Phi Y\|$