Optimal weight Choice of Frequentist Model Averaging Estimator

Model Averaging $\begin{cases} \text{Frequentist Model Averaging (FMA)} \\ \text{Bayesian Model Averaging (BMA)} \end{cases}$

Model Set-up: $Y = \underset{n\times k}{X}\beta + \underset{n\times m}{Z}\gamma + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I_n)$

focus regression auxiliary regression

1. unrestricted model (largest): $\hat\beta_u = \hat\beta_r - (X'X)^{-1}X'Z(Z'MZ)^{-1}Z'MY$, $\hat\gamma_u = (Z'MZ)^{-1}Z'MY$

   fully restricted model (only X): $\hat\beta_r = (X'X)^{-1}X'Y$, $M = I - X(X'X)^{-1}X'$

   Denote $\hat\theta = (Z'MZ)^{\frac{1}{2}}\hat\gamma_u$, $Q = (X'X)^{-1}X'Z(Z'MZ)^{-\frac{1}{2}}$, then $\hat\beta_u = \hat\beta_r - Q\hat\theta$

2. FMA estimator: $\hat\beta_f = \sum_{i=1}^{N} \lambda_i \hat\beta_{(i)}$, $\hat\mu_{(i)} = X\hat\beta_{(i)} + Z_i\hat\gamma_{(i)}$. $\hat\beta_{(i)}, \hat\gamma_{(i)}$ — i-th model

   weight: $\lambda_i = \lambda_i(\hat\theta, \hat\sigma^2)$, $\hat\sigma^2 = \frac{\|Y - X\hat\beta_u - Z\hat\gamma_u\|^2}{n}$

   λ_i should be based on $AIC_{(i)} = n\ln\frac{n-k-m}{n}\hat\sigma^2 + \frac{\hat\theta' P_i \hat\theta}{n}$, function of $\hat\sigma^2, \hat\theta$.

3. Let S_i — $m\times r_i$ selection matrix with rank $r_i \geq 0$ s.t. $S_i' = (I_{r_i} : 0)$

   New constrain: $S_i'^{*}\gamma = 0$, S_i gives variable disregarded,

   Denote $X_* = (X, Z)$, $\beta_*' = (\beta', \gamma')$, $R' = (0, S_i')$

   Task: Find β_* of $Y = X_*\beta + \varepsilon$ under $R'\beta_* = 0$, $\hat\beta_*' = (\hat\beta_{(i)}, \hat\gamma_{(i)})$

   Solve: $L = (Y - X_*\beta)'(Y - X_*\beta) + \lambda'(R'\beta_*) \Rightarrow \frac{\partial L}{\partial \beta_*} = -2X_*'Y + 2X_*'X_*\beta_* + R\lambda = 0$

   Solution: $\hat\gamma_{(i)} = (Z'MZ)^{-\frac{1}{2}}W_i\hat\theta$, $\hat\beta_{(i)} = \hat\beta_r - QW_i\hat\theta$

   $W_i = I - P_i$, $P_i = (Z'MZ)^{-\frac{1}{2}}S_i(S_i'(Z'MZ)^{-1}S_i)^{-1}S_i'(Z'MZ)^{-\frac{1}{2}}$

   $\Rightarrow \hat\beta_f = \sum_{i=1}^{N}\lambda_i\hat\beta_{(i)} = \hat\beta_r - QW\hat\theta$, $W = \sum_{i=1}^{N}\lambda_i W_i$

4. MSE: $E\|\hat\beta_f - \beta\|^2 = tr(E(\hat\beta_f - \beta)(\hat\beta_f - \beta)') = tr(\sigma^2 (X'X)^{-1} + Q E[(W\hat\theta - \theta)(W\hat\theta - \theta)']Q')$

(Reason: $\hat\beta_r \in span(X)$, $\hat\theta \in (span(X))^c$

$$E[(W\hat\theta - \theta)(W\hat\theta - \theta)'] = E[(W-I)\hat\theta\hat\theta'(W-I)] + \underbrace{E[(\hat\theta-\theta)(\hat\theta-\theta)']}_{\sigma^2 I_m}$$

$$+ E[(W-I)\hat\theta(\hat\theta-\theta)'] + \underline{E[(\hat\theta-\theta)\hat\theta'(W-I)]}$$

$$= \underbrace{\sum_{i=1}^{k} E[\lambda_i(\hat\theta,\hat\sigma^2)(\hat\theta-\theta)\hat\theta'](W_i - I)}$$

$$E[\lambda_i(\hat\theta,\hat\sigma^2)(\hat\theta-\theta)\hat\theta'] = E_{\hat\sigma^2}\{E_{\hat\theta}[\lambda_i(\hat\theta,\hat\sigma^2)(\hat\theta-\theta)\hat\theta']\}$$

$$= E_{\hat\sigma^2}\{\sigma^2 E_{\hat\theta}[\lambda_i(\hat\theta,\hat\sigma^2) I_m + \frac{\partial\lambda_i(\hat\theta,\hat\sigma^2)}{\partial\hat\theta}\hat\theta']\}$$ Stein's Lemma.

$$= \sigma^2 E[\lambda_i(\hat\theta,\hat\sigma^2) I_m + \frac{\partial\lambda_i(\hat\theta,\hat\sigma^2)}{\partial\hat\theta}\hat\theta']$$

Thus $E(\hat\beta_f-\beta)(\hat\beta_f-\beta)' = QE[(W\hat\theta-\hat\theta)(W\hat\theta-\hat\theta)']Q' - \hat\sigma^2 QQ' + \sigma^2(X'X)^{-1}$

$$+ \sigma^2 E[\psi_1(\hat\theta,\hat\sigma^2)] + \sigma^2 E[\psi_1'(\hat\theta,\hat\sigma^2)]$$

$$\psi_1(\hat\theta,\hat\sigma^2) = Q(W + \sum_{i=1}^{k}\frac{\partial\lambda_i(\hat\theta,\hat\sigma^2)}{\partial\hat\theta}\hat\theta W_i)Q'$$

Let $E\psi(\hat\theta,\hat\sigma^2) = \sigma^2 E\psi_1(\hat\theta,\hat\sigma^2) = E_{\hat\theta}[\sigma^2 E_{\hat\sigma^2}\psi_1(\hat\theta,\hat\sigma^2)]$

If $\Phi(\hat\theta,\hat\sigma^2)$ is differentiable,

$$E_{\hat\sigma^2}[\underbrace{\hat\sigma^2\Phi(\hat\theta,\hat\sigma^2)}_{\psi_1(\hat\theta,\hat\sigma^2)}] = \sigma^2 E_{\hat\sigma^2}[\Phi(\hat\theta,\hat\sigma^2) + \frac{2\hat\sigma^2}{n-k-m}\frac{\partial\Phi(\hat\theta,\hat\sigma^2)}{\partial\hat\sigma^2}]$$