L01 MSE-matrix criterion and MSE-criterion

1. MSCPE
   $\widehat{\theta} \in R^k$ is a point estimator for $\theta \in R^k$.

   (1) Sums and Cross Products from Error vector (SCPE)
   $\widehat{\theta} - \theta \in R^k$ is the error vector and $L(\theta, \widehat{\theta}) = (\widehat{\theta} - \theta)(\widehat{\theta} - \theta)' \in R^{k\times k}$ is a matrix-valued loss function for $\widehat{\theta}$. The components of this matrix are either Squares or Cross-Products of the components of the Error vector. So it is a SCPE matrix.

   (2) Mean of SCPE (MSCPE)
   $E(\widehat{\theta} - \theta) = \text{bias}(\widehat{\theta})$ and $R(\theta, \widehat{\theta}) = E[L(\theta, \widehat{\theta})] = E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)']$ is a matrix-valued risk function for $\widehat{\theta}$. We denoted it as MSCPE($\widehat{\theta}$) since it is the Mean of SCPE.
   But most books call it MSE-matrix for $\widehat{\theta}$ denoted as MSE-matrix($\widehat{\theta}$).

   (3) MSCPE-criterion
   For non-negative definite matrixes A and B, define
   $A - B \leq 0 \Longleftrightarrow B - A \geq 0 \Longleftrightarrow B - A$ is a non-negative definite matrix.

   $$\begin{aligned} \text{Estimator } \widehat{\theta} \text{ dominates estimator } \widetilde{\theta} &\Longleftrightarrow \text{MSCPE}(\widehat{\theta}) \leq \text{MSCPE}(\widetilde{\theta}) \text{ for all } \theta \\ &\Longleftrightarrow \text{MSE-matrix}(\widehat{\theta}) \leq \text{MSE-matrix}(\widetilde{\theta}) \text{ for all } \theta \\ &\Longleftrightarrow \widetilde{\theta} \text{ is inadmissible} \end{aligned}$$

   If $\widehat{\theta}$ is in a class and dominates all members in the class, then $\widehat{\theta}$ is the best estimator in that class by the MSCPE-criterion, or simply the MSE-matrix criterion.

   (4) MSCPE for unbiased estimators
   If $\widehat{\theta}$ is an unbiased estimator (UE) for θ, then MSCPE($\widehat{\theta}$) = MSE-matrix($\widehat{\theta}$) = Cov($\widehat{\theta}$).
   Here Cov($\widehat{\theta}$) is the variance-covariance matrix for $\widehat{\theta}$.
   Proof: With $\theta = E(\widehat{\theta})$, $\text{MSCPE}(\widehat{\theta}) = E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)'] = \text{Cov}(\widehat{\theta})$. $\Box$.

   (5) A relation
   $\text{MSE-matrix}(\widehat{\theta}) = \text{MSCPE}(\widehat{\theta}) = \text{Cov}(\widehat{\theta}) + [\text{bias}(\widehat{\theta})][\text{bias}(\widehat{\theta})]'$.

   $$\begin{aligned} \textbf{Proof } \text{MSCPE}(\widehat{\theta}) &= E\{[\widehat{\theta} - E(\widehat{\theta}) + E(\widehat{\theta}) - \theta][\widehat{\theta} - E(\widehat{\theta}) + E(\widehat{\theta}) - \theta]'\} \\ &= \text{Cov}(\widehat{\theta}) + [\text{bias}(\widehat{\theta})][\text{bias}(\widehat{\theta})]'. \end{aligned}$$

2. MSE
   $\widehat{\theta} \in R^k$ is a point estimator for $\theta \in R^k$.

   (1) SSE
   $\widehat{\theta} - \theta$ is the error, $L(\theta, \widehat{\theta}) = \|\widehat{\theta} - \theta\|^2 = \sum_{i=1}^k (\widehat{\theta}_i - \theta_i)^2$ is a non-negative valued loss function for $\widehat{\theta}$. We call it the SSE for $\widehat{\theta}$ since it is the Sum of Squared Errors.

   (2) MSE
   $E(\widehat{\theta} - \theta) = \text{bias}(\widehat{\theta})$, $R(\theta, \widehat{\theta}) = E[L(\theta, \widehat{\theta})] = E\|\widehat{\theta} - \theta\|^2$ is the corresponding risk function for $\widehat{\theta}$ denoted as MSE($\widehat{\theta}$) since it is the Mean of SSE.

   (3) MSE-criterion

   $$\begin{aligned} \text{Estimator } \widehat{\theta} \text{ dominates estimator } \widetilde{\theta} &\Longleftrightarrow \text{MSE}(\widehat{\theta}) \leq \text{MSE}(\widetilde{\theta}) \text{ for all } \theta \\ &\Longleftrightarrow \widetilde{\theta} \text{ is inadmissible} \end{aligned}$$

   If $\widehat{\theta}$ is in a class and dominates all members in the class, then $\widehat{\theta}$ is the best estimator in that class by the MSE-criterion.

(4) MSE for unbiased estimators

If $\widehat{\theta}$ is a unbiased estimator (UE) for θ, then $\mathrm{MSE}(\widehat{\theta}\,) = \sum_{i=1}^{k} \mathrm{var}(\widehat{\theta}_i)$ is the Total variance in $\widehat{\theta}$.

Proof: $\mathrm{MSE}(\widehat{\theta}\,) = E\|\widehat{\theta} - \theta\|^2 = E[\sum_i (\widehat{\theta}_i - \theta_i)^2 = \sum_i \mathrm{var}(\widehat{\theta}_i\,)$

(5) A relation

$\mathrm{MSE}(\widehat{\theta}\,) = \sum_{i=1}^{k} \mathrm{var}(\widehat{\theta}_i\,) + \|\mathrm{bias}(\widehat{\theta}\,)\|^2$.

Proof:
$$\begin{aligned} \mathrm{MSE}(\widehat{\theta}\,) &= E\|\widehat{\theta} - \theta\|^2 = E[\textstyle\sum_i (\widehat{\theta}_i - \theta_i)^2] = \sum_i E[\widehat{\theta}_i - E(\widehat{\theta}_i\,) + E(\widehat{\theta}_i\,) - \theta_i]^2 \\ &= \textstyle\sum_i E[\widehat{\theta}_i - E(\widehat{\theta}_i)]^2 + \sum_i [E(\widehat{\theta}_i) - \theta_i]^2 = \sum_i \mathrm{var}(\widehat{\theta}_i\,) + \|\mathrm{bais}(\widehat{\theta}\,)\|^2. \end{aligned}$$

3. Risk for $\widehat{\theta} \in R$.

$\widehat{\theta}$ is a point estimator for $\theta \in R$.

(1) $\text{MSE-matrix}(\widehat{\theta}\,) = \mathrm{MSE}(\widehat{\theta}\,) = E(\widehat{\theta} - \theta)^2$.

Proof: $\text{MSE-matrix}(\widehat{\theta}\,) = E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)']$. $\mathrm{MSE}(\widehat{\theta}\,) = E[(\widehat{\theta} - \theta)'(\widehat{\theta} - \theta)]$.

But $(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)' = (\widehat{\theta} - \theta)'(\widehat{\theta} - \theta) = (\widehat{\theta} - \theta)^2$.

So $\text{MSE-matrix}(\widehat{\theta}\,) = \mathrm{MSE}(\widehat{\theta}\,) = E(\widehat{\theta} - \theta)^2$. $\square$

(2) $\mathrm{MSE}(\widehat{\theta}\,) = \mathrm{var}(\widehat{\theta}\,) + [\mathrm{bias}(\widehat{\theta}\,)]^2$.

Proof:
$$\begin{aligned} \mathrm{MSE}(\widehat{\theta}\,) &= E(\widehat{\theta} - \theta)^2 = E[\widehat{\theta} - E(\widehat{\theta}\,) + E(\widehat{\theta}\,) - \theta]^2 \\ &= E[\widehat{\theta} - E(\widehat{\theta}\,)]^2 + [E(\widehat{\theta}\,) - \theta]^2 = \mathrm{var}(\widehat{\theta}\,) + [\mathrm{bais}(\widehat{\theta}\,)]^2. \end{aligned}$$

L02: MVUE and MTVUE

1. Risks
   When estimating θ by $\widehat{\theta}$, $E = (\widehat{\theta} - \theta)$ is the error and $E(\widehat{\theta} - \theta) = \text{bias}(\widehat{\theta})$.

   (1) $\text{MSE}(\widehat{\theta})$ for $\widehat{\theta} \in R$
   $(\widehat{\theta} - \theta)^2$, Squared Error (SE), is a loss when estimating θ by $\widehat{\theta}$.
   $E(\widehat{\theta} - \theta)^2$, the Mean of SE (MSE), is the risk denoted as $\text{MSE}(\widehat{\theta})$.

   (2) $\text{MSE}(\widehat{\theta})$ for $\widehat{\theta} \in R^k$
   $(\widehat{\theta} - \theta)'(\widehat{\theta} - \theta) = \sum_i (\widehat{\theta}_i - \theta_i)^2$, the Sum of Squared Errors (SSE), is a loss.
   $E[(\widehat{\theta} - \theta)'(\widehat{\theta} - \theta)]$, the Mean of SSE (MSE), is the risk denoted as $\text{MSE}(\widehat{\theta})$.
   This definition holds for all $k = 1, 2, ...$

   (3) MSE-matrix$(\widehat{\theta})$ for $\widehat{\theta} \in R^k$
   $(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)'$, a matrix containing Squares and Cross Products of the components of Error vector (SCPE), is a loss.
   $E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)']$ is the associated risk. For convenience it is denoted as MSE-matrix$(\widehat{\theta})$.
   When $k = 1$, MSE-matrix$(\widehat{\theta}) = \text{MSE}(\widehat{\theta})$.

2. Relations of risks and bias

   (1) $\theta \in R$
   $\text{MSE}(\widehat{\theta}) = \text{var}(\widehat{\theta}) + [\text{bias}(\widehat{\theta})]^2$.
   Proof:
   $$\begin{aligned}\text{MSE}(\widehat{\theta}) &= E(\widehat{\theta} - \theta)^2 = E[\widehat{\theta} - E(\widehat{\theta}) + E(\widehat{\theta}) - \theta]^2 \\ &= E[\widehat{\theta} - E(\widehat{\theta})]^2 + [E(\widehat{\theta}) - \theta]^2 = \text{var}(\widehat{\theta}) + [\text{bias}(\widehat{\theta})]^2.\end{aligned}$$

   (2) $\theta \in R^k$
   $\text{MSE}(\widehat{\theta}) = \sum_i \text{var}(\widehat{\theta}_i) + \|\text{bias}(\widehat{\theta})\|^2$.
   Proof:
   $$\begin{aligned}\text{MSE}(\widehat{\theta}) &= E[(\widehat{\theta} - \theta)'(\widehat{\theta} - \theta)] = E[\textstyle\sum_i (\widehat{\theta}_i - \theta_i)^2] \\ &= \textstyle\sum_i E[\widehat{\theta}_i - E(\widehat{\theta}_i) + E(\widehat{\theta}_i) - \theta_i]^2 \\ &= \textstyle\sum_i E[\widehat{\theta}_i - E(\widehat{\theta}_i)]^2 + \sum_i [E(\widehat{\theta}_i) - \theta_i]^2 = \sum_i \text{var}(\widehat{\theta}_i) + \|\text{bias}(\widehat{\theta})\|^2.\end{aligned}$$

   (3) $\theta \in R^k$
   MSE-matrix$(\widehat{\theta}) = \text{Cov}(\widehat{\theta},) + [\text{bias}(\widehat{\theta})][\text{bias}(\widehat{\theta})]'$.
   Proof:
   $$\begin{aligned}\text{MSE-matrix}(\widehat{\theta}) &= E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)'] \\ &= E[\widehat{\theta} - E(\widehat{\theta}) + E(\widehat{\theta}) - \theta)][\widehat{\theta} - E(\widehat{\theta}) + E(\widehat{\theta}) - \theta)]' \\ &= E[\widehat{\theta} - E(\widehat{\theta})][\widehat{\theta} - E(\widehat{\theta})]' + [E(\widehat{\theta}) - \theta)][E(\widehat{\theta}) - \theta)]' \\ &= \text{Cov}(\widehat{\theta}) + [\text{bias}(\widehat{\theta})][\text{bias}(\widehat{\theta})]'.\end{aligned}$$

   (4) Risk for U-estimators
   Suppose $\widehat{\theta}$ is a U-estimator. Then $\text{MSE}(\widehat{\theta}) = \text{var}(\widehat{\theta})$ if $\widehat{\theta} \in R$.
   $\text{MSE}(\widehat{\theta}) = \sum_i \text{var}(\widehat{\theta})$, the Total Variance in $\widehat{\theta}$, and MSE-matrix$(\widehat{\theta}) = \text{Cov}(\widehat{\theta})$.
   Ex1: $\overline{x}$ and s^2 are sample mean and sample variance of a sample of size n from $N(\mu, \sigma^2)$.
   Suppose $\theta = \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}$ is estimated by $\widehat{\theta} = \begin{pmatrix}\overline{x} \\ s^2\end{pmatrix}$. Find $\text{MSE}(\widehat{\theta})$.
   $\overline{x}$ is an UE for μ and s^2 is an UE for σ^2. So $\text{MSE}(\widehat{\theta}) = \text{var}(\overline{x}) + \text{var}(s^2)$.
   But $\overline{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) = \text{Gamma}(\alpha, \beta)$ with $\alpha = \frac{n-1}{2}$ and $\beta = 2$. From $\text{Gamma}(\alpha, \beta) \sim (\alpha\beta, \alpha\beta^2)$, $\text{var}\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{n-1}{2} \cdot 2^2$, i.e., $\frac{(n-1)^2}{\sigma^4}\text{var}(s^2) = 2(n-1)$. Thus $\text{var}(s^2) = \frac{2\sigma^4}{n-1}$. Therefore $\text{MSE}(\widehat{\theta}) = \frac{\sigma^2}{n} + \frac{2\sigma^4}{n-1}$.

Ex2: For θ and $\widehat{\theta}$ in Ex1, find MSE-matrix($\widehat{\theta}$).

Because $\widehat{\theta}$ is an UE for θ, MSE-matrix($\widehat{\theta}$) = Cov($\widehat{\theta}$). But the distributions of $\overline{x}$ and s^2 are independent. So MSE-matrix($\widehat{\theta}$) $= \begin{pmatrix} \mathrm{var}(\overline{x}) & 0 \\ 0 & \mathrm{var}(s^2) \end{pmatrix} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{pmatrix}$.

3. MVUE and MTVUE

Let UE(θ) be the collection of all unbiased estimator for θ

(1) MVUE for $\theta \in R$

$\widehat{\theta}$ is the best estimator in UE(θ) by MSE

$\iff$ $\widehat{\theta} \in$ UE(θ) and $\mathrm{var}(\widehat{\theta}) \le \mathrm{var}(\widetilde{\theta})$ for all $\widetilde{\theta} \in$ UE(θ)

$\iff$ $\widehat{\theta}$ is the Minimum Variance Unbiased Estimator (MVUE) for θ.

(2) MVUE for $\theta \in R^k$

$\widehat{\theta}$ is the best estimator in UE(θ) by MSE-matrix

$\iff$ $\widehat{\theta} \in$ UE(θ) and $\mathrm{Cov}(\widehat{\theta}) \le \mathrm{Cov}(\widetilde{\theta})$ for all $\widetilde{\theta} \in$ UE(θ)

$\iff$ $\widehat{\theta}$ is the Minimum Variance-covariance matrix Unbiased Estimator, MVUE for θ.

(3) MTVUE for $\theta \in R^k$

$\widehat{\theta}$ is the best estimator in UE(θ) by MSE

$\iff$ $\widehat{\theta} \in$ UE(θ) and $\sum_i \mathrm{var}(\widehat{\theta}_i) \le \sum_i \mathrm{var}(\widetilde{\theta})$ for all $\widetilde{\theta} \in$ UE(θ)

$\iff$ $\widehat{\theta}$ is the Minimum Total Variance Unbiased Estimator (MTVUE) for θ.

(4) Relations

If $\widehat{\theta}$ is a MVUE for θ, then $\widehat{\theta}$ is a MTVUE for θ.

Suppose $\widehat{\theta}$ is MTVUE for θ. If MVUE exists, then MTVUE and MVUE are equal.

Ex3: For $A \ge 0$ and $B \ge 0$, $\mathrm{tr}(A) \le \mathrm{tr}(B)$ does not imply $A \le B$.

$A = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}$ give a counterexample.