L01 Estimators and predictors

1. Estimator classes and predictor classes

$\theta \in R^k$ is a parameter vector from a population system. $y_1, ..., y_n$ is a random sample from the system and $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. $y_* \in R^k$ is an unknown random vector. Suppose $\widehat{\theta} \in R^k$ and $\widehat{y}_* \in R^k$ are statistics.

(1) Estimator classes
When θ is estimated by $\widehat{\theta}$, $\widehat{\theta}$ is called a point estimator (PE) for θ. If $E_\theta(\widehat{\theta} - \theta) \equiv 0$, i.e., $E_\theta(\widehat{\theta}) \equiv \theta$, then $\widehat{\theta}$ is an unbiased estimator (UE) for θ. The collection of all UEs for θ is denoted as UE(θ).
$\widehat{\theta}$ is a linear estimator (LE) if $\widehat{\theta} = Ly$ for some L. The collection of all linear unbiased estimators (LUEs) for θ is denoted as LUE(θ). Clearly

$$\text{LUE}(\theta) \subset \text{UE}(\theta).$$

(2) Predictor classes
When y_* is predicted by $\widehat{y}_*$, $\widehat{y}_*$ is called a point predictor (PP) for y_*. If $E_\theta(\widehat{y}_* - y_*) \equiv 0$, i.e., $E_\theta(\widehat{y}_*) \equiv E_\theta(y_*)$, then $\widehat{y}_*$ is an unbiased predictor (UP) for y_*. The collection of all UP for y_* is denoted as UP(y_*).
$\widehat{y}_*$ is a linear predictor (LP) if $\widehat{y}_* = Ly$ for some L. The collection of all linear unbiased predictors (LUPs) for y_* is denoted as LUP(y_*). Clearly

$$\text{LUP}(y_*) \subset \text{UP}(y_*).$$

(3) Same classes

$$\text{UP}(y_*) = \text{UE}(E(y_*)) \qquad \text{and} \qquad \text{LUP}(y_*) = \text{LUE}(E(y_*)).$$

Proof. Show the first one only.

$$\begin{aligned} \widehat{y}_* \in \text{UP}(y_*) &\iff E_\theta(\widehat{y}_* - y_*) \equiv 0 \iff E_\theta(\widehat{y}_*) \equiv E_\theta(y_*) \iff E_\theta[\widehat{y}_* - E_\theta(y_*)] \equiv 0 \\ &\iff \widehat{y}_* \in \text{UE}(E(y_*)). \end{aligned}$$

2. Comparing estimators and predictors

(1) Risk function for estimators
With matrix-valued loss $L(\widehat{\theta}, \theta)$ when θ is estimated by $\widehat{\theta}$, the risk is $R_{\widehat{\theta}}(\theta) = E_\theta[L(\widehat{\theta}, \theta)]$.
If both $\widehat{\theta}$ and $\widetilde{\theta}$ are estimators for θ and $R_{\widehat{\theta}}(\theta) \le R_{\widetilde{\theta}}(\theta)$ for all θ, i.e., $R_{\widetilde{\theta}}(\theta) - R_{\widehat{\theta}}(\theta)$ is a non-negative definite matrix, then $\widetilde{\theta}$ is inadmissible since it is dominated by $\widehat{\theta}$.
If in a class of estimators for θ, $\widehat{\theta}$ dominates all other estimators, then $\widehat{\theta}$ is the best estimator in that class w.r.t. the specified risk.

(2) Risk function for predictors
With matrix-valued loss $L(\widehat{y}_*\, y_*)$ when y_* is predicted by $\widehat{y}_*$, $R_{\widehat{y}_*}(\theta) = E_\theta[L(\widehat{y}_*, y_*)]$ is the risk.

If both $\widehat{y}_*$ and $\widetilde{y}_*$ are predictors for y_* and $R_{\widehat{y}_*}(\theta) \leq R_{\widetilde{y}_*}(\theta)$ for all θ, i.e., $R_{\widetilde{y}_*}(\theta) - R_{\widehat{y}_*}(\theta)$ is a non-negative definite matrix, then $\widetilde{y}_*$ is inadmissible since it is dominated by $\widehat{y}_*$.
If in a class of predictors for y_*, $\widehat{y}_*$ dominates all other predictors, then $\widehat{y}_*$ is the best estimator in that class w.r.t. the specified risk.

(3) Same classes
To identify best predictor in UP(y_*) we need a risk function. To identify best estimator in UE$(E(y_*))$ we need a risk function. But two statistics classes UP(y_*) and UE$(E(y_*))$ are equal.
To identify best predictor in LUP(y_*) we need a risk function. To identify best estimator in LUE$(E(y_*))$ we need a risk function. But two statistics classes LUP(y_*) and $LUE(E(y_*))$ are equal.

3. Two risk functions

(1) Two risk functions
For estimation,
$E[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'] \in R^{k\times k}$ is MSCPE risk and $E[(\widehat{\theta}-\theta)'(\widehat{\theta}-\theta)] \in R$ is MSE risk.
If $\widehat{\theta}$ dominates $\widetilde{\theta}$ by MSCPE risk, then the domination holds by MSE risk.
For prediction,
$E[(\widehat{y}_*-y_*)(\widehat{y}_*-y_*)'] \in R^{k\times k}$ is MSCPE risk and $E[(\widehat{y}_*-y_*)'(\widehat{y}_*-y_*)] \in R$ is MSE risk.
If $\widehat{y}_*$ dominates $\widetilde{y}_*$ by MSCPE risk, the the domination holds by MSE risk.

Ex1: If $\widehat{\theta}$ is the best estimator for θ in a class by MSCPE, then it is also the best in the class by MSE. If $\widehat{y}_*$ is the best predictor for y_* in a class by MSCPE, then it is also the best in the class by MSE.

(2) Best estimator in UE(θ) and best predictor in UP(y_*) by MSCPE
If $\widehat{\theta} \in$ UE(θ), then MSCPE risk is $E[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'] = \text{Cov}(\widehat{\theta})$.
Thus the best estimator in UE(θ) is the minimum variance-covariance matrix unbiased estimator (Minimum V-C UE).
If $\widehat{y}_* \in$ UP(y_*), then MSCPE risk is $E[(\widehat{y}_*-y_*)(\widehat{y}_*-y_*)'] = \text{Cov}(\widehat{y}_*-y_*)$.

Ex2: If $\widehat{\theta} \in$ UE(θ), then MSE risk is $E[(\widehat{\theta}-\theta)'(\widehat{\theta}-\theta)] = \sum_i \text{var}(\widehat{\theta}_i)$.
So the best estimator in UE(θ) by MSE risk is minimum total variance unbiased estimator.

(3) The case where y and y_* are independent
If y_* is independent to y, then the best predictor for y_* in UP(y_*) by MSCPE and the best estimator for $E(y_*)$ in UE$(E(y_*))$ by MSCPE are equal.

Proof.
$$
\begin{aligned}
& \widehat{y}_* \text{ is the best predictor in UP}(y_*) \text{ by MSCPE} \\
\iff & \widehat{y}_* \in \text{UP}(y_*) \text{ and } \text{Cov}(\widehat{y}_*-y_*) \leq \text{Cov}(\widetilde{y}_*-y_*) \text{ for all } \widetilde{y}_* \in \text{UP}(y_*) \\
\iff & \widehat{y}_* \in \text{UE}(E(y_*)) \text{ and } \text{Cov}(\widehat{y}_*)+\text{Cov}(y_*) \leq \text{Cov}(\widetilde{y}_*)+\text{Cov}(y_*) \\
& \quad \text{for all } \widetilde{y}_* \in \text{UE}(E(y_*)) \\
\iff & \widehat{y}_* \in \text{UE}(E(y_*)) \text{ and } \text{Cov}(\widehat{y}_*) \leq \text{Cov}(\widetilde{y}_*) \text{ for all } \widetilde{y}_* \in \text{UE}(E(y_*)) \\
\iff & \widehat{y}_* \text{ is the best estimator for } E(y_*) \text{ in UE}(E(y_*)) \text{ by MSCPE.}
\end{aligned}
$$

(4) Two sufficient conditions
(i) If $\widehat{\theta} \in$ UE(θ), $\widehat{\theta} = f(S)$ and S is a sufficient and complete statistics, then $\widehat{\theta}$ is the best estimator in UE(θ) by MSCPE.
(ii) If $\widehat{\theta} \in$ UE(θ) and $\text{Cov}(\widehat{\theta}) = \text{CRLB}(\theta) = [nI(\theta)]^{-1}$, then $\widehat{\theta}$ is the best estimator in UE(θ) by MSCPE.