Data-driven calibration of linear estimators with minimal penalties (Sylvain Arlot and Francis Bach)

Goals:
choosing the kernel in multiple kernel Learning (MKL)
choosing the regularization parameter in kernel ridge regression
$\Leftrightarrow$ **Select from multiple linear estimators** $\widehat{Y} = A_\lambda Y$, $\lambda \in \Lambda$

Penalization approach: $\boxed{\widehat{\lambda} \in \arg\min_{\lambda \in \Lambda} \{ \| Y - A_\lambda Y \|^2 + \mathrm{pen}(A_\lambda) \}}$

Ideal penalty $\mathrm{pen}(A_\lambda) = \sigma^2 \mathrm{tr}(A_\lambda^2)$ leads to optimal selection, but depends on unknown noise variance σ^2

Minimal penalty $\mathrm{pen}(A_\lambda) = C[2\mathrm{tr}(A_\lambda) - \mathrm{tr}(A_\lambda^2)]$ leads to a sharp jump in $\mathrm{tr} A_{\hat{\lambda}(C)}$ around $C = \sigma^2$

Allows **data-driven** estimation of σ^2 and non-asymptotic **oracle inequalities**