Threshold Stability Across Folds
(lower CV = more deployment-stable)