BOX A.6: LEAVE-ONE-EARNINGS-OUT (LOEO)

LOEO provides earnings-respectful validation by holding out a single earnings window and evaluating whether a model trained on all remaining days (*including past earnings windows*) can accurately predict targets in that unseen event window. Earnings anchors (the date where earning _date == 0) were identified using the "days-since-earnings" counters on the wide event-engineered dataset for the ticker (e.g., wide_google_aug_tied). Around each anchor (A) a test window of ±W business days is defined, centred on the event day (when *W=2 yields five test days (A−2,…,A+2); W=5 yields eleven; W=10 yields twenty-one*). To prevent look-ahead from the prediction horizon, an embargo of H trading days is applied on both sides of that test window. In practical terms, the training set for that fold consists of every date outside the window and the embargo [$A−W−H$, $A+W+H$]; the test set is exactly the ±W block.
Similar to the walk-forward CV regime, performance is computed per earnings fold and then aggregated, using the same three evaluation metrics: hit rate, Spearman's ρ, and conditional strategy return. As before, reporting both the cross-fold mean, and the last (most recent) fold distinguishes long-run robustness from current-regime performance.
LOEO complements walk-forward CV. Combined, they couple robustness across market regimes with evidence of generalization to the next earnings event, providing a complete and realistic read on deployable performance.

Risks & Limitations:
1. *Early anchors may be dropped at alignment*: when initial earnings dates fall within the warm-up window used for rolling standardization and safeguards, LOEO excludes those folds by design. This reduces the number of evaluable events. Mitigations include extending the historical span so ~8–12 earnings events remain after alignment; in this project, at least nine folds were evaluated per company.
2. *Tight LOEO windows yield small test samples*: with few observations per fold, metrics such as hit rate and Spearman's ρ can fluctuate. In this setting, RidgeCV often remains competitive (lower variance helps when data per fold are scarce), whereas XGB can exhibit fold-to-fold volatility in hit/ρ. Mitigations include using modest (but not tiny) windows (e.g., ±5 or ±10), reporting confidence intervals across folds, and complementing LOEO with walk-forward CV to show performance outside the earnings window. Around earnings, the hybrid XGB approach particularly benefits from Ridge-based top K pre-selection, which serves as a valuable safeguard.