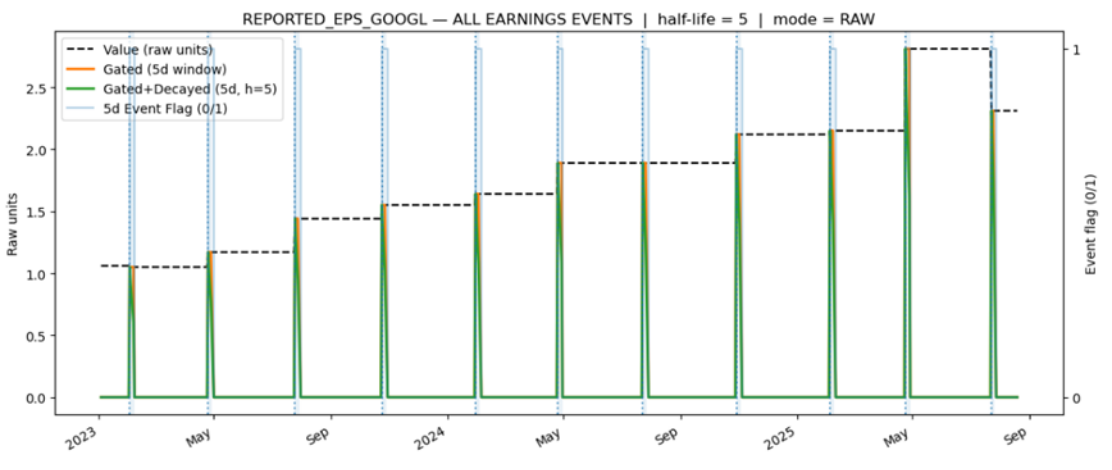
BOX A.3: EVENT-AWARE FEATURE ENGINEERING

Macroeconomic variables (CPI, Trade, Unemployment, GDP, DSR, FOMC) and corporate earnings (EPS, revenue, surprises) are low-frequency events. If values are simply forward filled, the model sees the same number repeated for weeks or months. This creates two risks: (A) *Spurious trend learning* – the model might misuse a constant value as a proxy for time/trend; (B) *Dilution of impact* – the real market effect of a release is concentrated around the event, not uniformly spread until the next update. **To mitigate these risks and to properly align values with their true market impact, event-aware features were engineered**:

- **Event flags (0/1):** Mark days inside an event window (e.g. 1-day, 2-day, 5-day after release). *Example: earnings_date_5d_googl.*
- **Gated features:** Multiply the raw or standardized value by the event flag. *Example: reported_eps_googl__gate_earnings_date_5d_googl.*
- **Decayed features:** Multiply the raw value by a decay weight that halves every h days. *Example: reported_eps_googl__decay_h5.*
- **Gated+Decayed features:** Combine both. *Example: reported_eps_googl__gate_earnings_date_5d_googl_dec_h5.*

The process was applied to all macroeconomic features with half-lives tuned to horizon T+10 (7–20 days); and to earnings for each ticker with half-life = 5 days. Features were built for both raw and standardized (_gz) values. Raw columns were preserved (drop_raw=False) to avoid information loss and keep interpretability. Visualizations were produced to confirm the design and improve interpretability.

Example of Engineered Features — Raw EPS vs. Gated and Gated + Decayed Windows (GOOGL)



Contribution to Modelling: Event flags together with gated+decayed features eliminated misleading persistence, prevented the model from assuming that an EPS or CPI value should influence returns months later, and directed learning toward true event responses—capturing event-driven spikes in explanatory variables, reflecting the diminishing impact of releases, and avoiding overweighting of prolonged flat periods. The design allowed different horizons: 1d/2d gates – for short-term effects (T+2, T+5); and 5d gated+decayed – for medium-term effects and predictions (T+10). An EDA check confirmed that the new gated+decayed features show a significant correlation with T+10 returns. Finally, by transforming static low-frequency values into time-aware, event-sensitive features, modelling produced cleaner signals, reduced spurious trend dependence, and enhanced short- to mid-term predictive power.