Short answer: your slate is strong, but there are high-leverage gaps. Below are the biggest additions I'd make—each chosen to boost *out-of-sample* predictability for multi-horizon bqx across all 8 pairs while staying leakage-safe.

Quick gap scan (what's missing)

1. **Structure-aware FX math** beyond triangles (cointegration / ECM).
2. **Regime/change detection** (statistical, not just heuristics).
3. **Spectral/shape features** (frequency bands, wavelets, SSA).
4. **Realized-volatility family** (range-based, bipower, quarticity).
5. **Impact/liquidity proxies** beyond spread/volume (Kyle/Amihud style).
6. **Cross-sectional signals** (ranks/dispersion/relative value) baked into features.
7. **Learned representations** (panel autoencoders / contrastive encoders).
8. **Target & loss shaping** for multi-horizon coherence and uncertainty.

What to add (and why)

A) FX structure & error-correction (high impact)

- **Johansen cointegration** on each USD triangle and key clusters (EUR–GBP–USD, AUD–NZD–USD).
  - Add: cointegration vectors, **error-correction term (ECT)**, and ΔECT.
  - Why: ECT mean-reverts; its level/velocity is predictive for 15–75 min horizons.
- **Cross-pair regression residuals with parabolic terms** (you have triangles; extend to *sister pairs*):
  - Add residual level/z, residual curvature (\tilde a_2) and slope (\tilde a_1).
  - Why: highlights transient mispricings your bqx will often "catch up" to.

B) Robust realized-volatility set (cheap, powerful)

Compute on 1m and 5m bars (up to i only):

- **Parkinson**, **Garman–Klass**, **Rogers–Satchell**, **Yang–Zhang** vol.
- **Bipower variation** and **realized quarticity** (proxy for jump risk).
- **Vol-of-vol** and **vol acceleration** (ΔEWMAvol, second difference).
  Why: bqx predictability is regime-dependent; these quantify regime cleanly and are orthogonal to your current std/EWMA.

C) Regime & change-point detection

- **HMM/Markov regimes** on returns or vol (K=3: calm/trend/shock). Use filtered state probs (P(state=k)_i).
- **Bayesian Online Change Point Detection (BOCPD)** run-length & hazard features.
- **CUSUM statistics** on returns and on your (\tilde a_2) curvature stream.
  Why: sharp improvements in forecasts when models know if the distribution just shifted.

D) Spectral / shape analytics (multi-scale, orthogonal)

- **FFT band energies** (very-short 2–4m, short 5–15m, medium 20–60m) on bqx and rate; include band-energy ratios and dominant frequency.
- **Wavelet packet energy** (e.g., Daubechies 4) at 3–4 levels; add **cross-window coherence** flags.
- **Singular Spectrum Analysis (SSA)**: first 2–3 reconstructed components' slopes/curvatures.
  Why: picks up cyclic build-ups and microstructure rhythms that precede breakouts/reversions.

E) Microstructure impact proxies (if you have basic volume/tick)

- **Amihud illiquidity**: |ret| / volume (1m, 5m).
- **Kyle's λ surrogate**: slope of Δprice on signed volume (rolling OLS).
- **VPIN-lite**: imbalance of up- vs down-volume (you already have uptick/downtick—aggregate to a toxicity score).
- **Spread*vol** interactions and **book-pressure proxy**: spread z × volume z.
  Why: future bqx (a momentum derivative) is especially sensitive to flow/impact regimes.

F) Cross-sectional (8-pair panel) features

- **Cross-sectional ranks/percentiles** of: return, bqx, (\tilde a_1), (\tilde a_2), vol, ECT.
- **Dispersion** metrics (std across 8 pairs) for those same quantities.
- **Breadth**: fraction of pairs with positive (\tilde a_2) (acceleration breadth) and positive (\tilde a_1) (trend breadth).
  Why: panel context (synchrony vs rotation) is a leading indicator for continuation odds.

G) Multi-resolution: add 30m & 60m layers

You have 1m/5m/15m. Add:

- 30m/60m **vol, trend slope, (\tilde a_2)**, and **alignment flags** (e.g., sign agreement of (\tilde a_2) across 5→15→30→60).
  Why: horizons up to +75m benefit from information carried at 30–60m.

H) Dual-domain *derivative* comparisons (rate ↔ bqx)

You compare levels; add **differential dynamics**:

- **Gap velocity/acceleration**: Δ((\tilde a_1^{rate}-\tilde a_1^{bqx})), Δ((\tilde a_2^{rate}-\tilde a_2^{bqx})).
- **Apex timing spread**: (t^_*{rate}* - *t*^_{bqx}) from parabola.
  Why: the *speed of decoupling/recoupling* often predicts the next 15–45 minutes.

I) Learned representations (panel-wise, leakage-safe)

- **Autoencoder on feature blocks** (per timestamp × 8 pairs): keep the bottleneck (e.g., 8–16 dims) as **system embeddings**.
- **Temporal contrastive learning** (TS-TCC/SimCLR-style) on bqx panels; use embeddings as features.
  Why: compresses nonlinear cross-pair structure and denoises hand-engineered redundancy.

Target & loss shaping (often bigger lift than more features)

- **Vol-scaled Δ targets** (you already plan this) + **quantile heads** (τ=.1/.5/.9) to get calibrated uncertainty.
- **Multi-horizon coherence**: add penalty if predicted |Δbqx| is *not* non-decreasing with horizon (soft monotonicity).
- **Directional auxiliary**: BCE on sign(Δbqx) per horizon to sharpen turns.
- **Horizon embeddings** (if using a shared model) so the encoder learns horizon-specific patterns.

Data-health & stability features (cheap insurance)

- **Missing-rate**, **FFILL-count**, **stale-tick flag** per pair.
- **Feature drift z-scores** vs trailing 24h means for key blocks (vol, spread, (\tilde a_2)).
  Why: models behave better when they *see* data quality/regime drift.

What to ship first (12 additions with best ROI)

1. **ECT (error-correction term)** per triangle + ΔECT.
2. **Garman–Klass & bipower variation** (1m & 5m).
3. **Vol-of-vol** and **vol acceleration**.
4. **HMM state probabilities** (K=3) on returns.
5. **BOCPD run-length** on returns.
6. **FFT band energies** (short/medium) on bqx; band-energy ratios.
7. **Wavelet energy** level-1/2 on bqx.
8. **Amihud illiquidity** and **Kyle λ proxy**.
9. **Cross-sectional ranks** of (\tilde a_2), (\tilde a_1), and vol across 8 pairs.
10. **Apex timing spread** (t^_*{rate}* - *t*^_{bqx}).
11. **30m/60m layer**: vol, (\tilde a_2), alignment flags.
12. **Autoencoder bottleneck** (8–16 dims) from the full panel at time i.

Guardrails & pruning

- Keep all windows **ending at i**; no peeking.
- Winsorize extreme z-scores (e.g., at 4–5σ) to stabilize training.
- Expect multicollinearity—**groupwise PCA/PLS** per block (volatility, parabolic, spectral) and keep a few principal components + a handful of interpretable "anchor" features.
- Run **rolling feature importance** (e.g., time-slice SHAP) to prevent over-fitting to a single regime.
