Interval Forecasting of Cryptocurrency Returns using Quantile Regression Forests: An Application to the Solana Ecosystem

Quantile Regression Forests vs Linear Quantile Regression and LightGBM

James Lewis

Table of contents

1 Interval Forecasting of Cryptocurrency Returns using Quantile Regression Forests: An Application to the Solana Ecosystem

Quantile Regression Forests vs Linear Quantile Regression and LightGBM

Abstract. (200–300 words placeholder) Problem, data (12-h bars; 72-h target), models (QRF, LQR, LGBM), rolling CV (120/24/6), metrics (pinball, coverage, width), key results + trading relevance.

Keywords: Quantile Regression Forests; Conformal prediction; Crypto markets; Forecast intervals.

2 1 Introduction

2.1 1.1 Motivation

Risk management in crypto markets depends on the **distribution** of returns, not only their central tendency. Mid-cap Solana tokens exhibit heavy tails, skewness and regime dependence; under such conditions, symmetric error bands around point forecasts are misleading. Interval forecasts, expressed as conditional quantiles $q_\tau(x_t)$, are directly decision-relevant: they inform position sizing, stop placement and drawdown control, and they admit explicit calibration tests (e.g., whether an 80 % band achieves approximately 80 % empirical coverage). The appropriate scoring rule is the **pinball loss**,

$$L_\tau(y, \hat{q}_\tau) \;=\; (\tau - \mathbf{1}\{y < \hat{q}_\tau\})\,(y - \hat{q}_\tau),$$

a strictly proper objective that rewards calibrated asymmetry rather than squared error [add citation]. In this project, calibration and **sharpness** (narrow intervals at target coverage) are the primary goals.

2.2 1.2 Problem and scope

The task is to forecast **72-hour forward log returns** for **mid-cap Solana tokens** on a **12-hour** cadence. Let

$$\text{return\_72h}_t \;=\; \log P_{t+72\text{h}} - \log P_t,$$

constructed as a six-step forward difference on the 12-hour grid; horizons therefore overlap. The modelling sample runs from **5 December 2024 00:00** to **3 June 2025 00:00**. The universe comprises **23** tokens satisfying market-capitalisation ($30 m) and listing-age (3 months) filters. Because several names are crude or meme-styled, tickers are not enumerated here; a complete list is given in *Methods* and the *Appendix* (with neutral labels where appropriate).

All inputs are aligned on the 12-hour grid with preserved **imputation masks** to track data quality through rolling evaluation. Features span: momentum (1/3-bar returns, RSI/PROC/Stochastic oscillators), volatility (realised volatility, ATR, Bollinger bandwidth), liquidity and microstructure (bid-ask spread, top-of-book depth, volume), on-chain activity (changes in unique wallets/holders, transaction counts), and cross-asset context (SOL beta and level changes). Leakage is controlled via forward shifts and rolling windows; full definitions appear in *Data & Features*.

2.3 1.3 Approach and contributions

We estimate a grid of conditional quantiles $\tau \in \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$ using three model classes held to a **common feature set** and a **blocked rolling** design (*train 120 / calibrate 24 / test 6*, step 6) to mimic deployment and preserve temporal order.

Our **core** model is **Quantile Regression Forests (QRF)**, a non-parametric ensemble that estimates conditional quantiles without distributional assumptions [Meinshausen (2006)]. Two baselines are included for a fair comparison: **Linear Quantile Regression (LQR)** [add citation], providing a parametric benchmark, and **LightGBM in quantile mode** [add citation]. Because empirical coverage matters operationally, we apply **split-conformal** calibration to LightGBM's lower/upper quantiles to achieve nominal coverage under mild conditions [add citation]. Additional calibration refinements (non-crossing adjustments, median bias correction, regime-aware offsets) are described and justified in *Methods*; they are not relied upon here.

Evaluation focuses on **pinball loss** by τ, **empirical coverage** for 80 % (q10–q90) and 90 % (q05–q95) bands, and **interval width**. To assess statistical significance, we use the **Diebold–Mariano** test on pinball-loss differentials [add citation]. Preliminary results indicate that QRF improves tail accuracy relative to LightGBM; for example, at $\tau = 0.10$ the mean pinball is **0.1245 (QRF)** versus **0.1571 (LightGBM)**—a **20.8 %** reduction (**provisional; to be confirmed**) (Table (**?**):pinball-by-tau, [PATH]).

Taken together, the contribution is a calibrated, reproducible interval-forecasting pipeline for Solana mid-caps: a distribution-native method (QRF) compared against parametric and boosting baselines under a rigorous rolling protocol, with performance reported in terms that matter to trading—coverage and sharpness, not only central accuracy.

2.4 1.4 Research questions and hypotheses

This study addresses four questions:

1. **Predictive accuracy.** Do QRFs achieve **lower pinball loss** across τ than LQR and LightGBM under the rolling 120/24/6 protocol for 72-hour returns? **Hypothesis:** Yes, especially at the tails, where non-linear, heteroskedastic structure dominates [add citation].
2. **Calibrated sharpness.** At nominal 80 %/90 % coverage, which method yields the **narrowest calibrated** intervals? **Hypothesis:** LightGBM (post-conformal) tends to **over-cover** via wider bands; QRF attains **sharper tails** at target coverage; LQR **under-covers** away from the median [add citation].
3. **Robustness.** Are the conclusions stable to reasonable changes in rolling design and calibration choices (e.g., half-life, regime thresholds, non-crossing enforcement)? **Hypothesis:** Qualitative conclusions are stable; non-crossing reduces pathological crossings with negligible loss penalty [add citation].
4. **Drivers of width and asymmetry.** Which feature groups widen or narrow intervals? **Hypothesis:** Volatility and liquidity features dominate width; SOL cross-asset deltas shift the median; on-chain growth widens the right tail in influx regimes.

3 Literature review

- Interval / quantile forecasting and proper scoring.
- QRF (non-parametric), LQR (parametric), boosting (LightGBM).
- Conformal prediction / CQR; calibration in finite samples.
- Crypto specifics: heavy tails, regime shifts, microstructure & on-chain features.

  Figure refs to add later: reliability diagrams, schematic of interval scoring (optional).

4 Data and feature engineering

4.1 Universe and period

- Tokens ($50m mcap; 3 months listed), 12-h bars; target `return_72h`.

4.2 Cleaning, alignment, imputation

- Timestamp alignment; missing bin handling; imputation masks.

4.3 Feature set

- **Momentum** (logret_12h/36h, RSI/PROC/Stoch).
- **Volatility** (realised vol, ATR, Parkinson).
- **Liquidity/microstructure** (spread, depth, OBV, price_volume).
- **On-chain** (Δ wallets, tx counts).
- **Cross-asset** (SOL, ETH).

4.4 EDA highlights

- Return skew/heavy tails; regime markers; missingness summary.

5 Methods

5.1 Rolling evaluation

- Blocked per-token: train 120, cal 24, test 6; step 6.

5.2 Models

5.2.1 QRF (core)

- quantile_forest.RandomForestQuantileRegressor; tuned (Optuna); decay weights; $\{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$
- Calibration: residual quantile offsets by regime; median bias correction; isotonic non-crossing.

5.2.2 Baselines

- LQR (statsmodels) same grid; non-crossing.
- LightGBM quantile; conformal for 80% interval + interpolation.

5.3 Metrics & tests

- Pinball loss; empirical coverage + width; Diebold–Mariano tests; optional WIS.

5.4 Key equations

- Pinball loss, DM statistic (list equations here with labels eq:pinball, eq:dm).

6 Results

6.1 Global comparison

6.2 Visual overlays (representative tokens)

Interpretation

QRF sharp tails + good median; LightGBM wider (over-coverage); LQR under-coverage.

Pointers to regime-conditional results (next chapter).

7 Trading application

- Sizing rule from quantiles (describe formula).
- Backtest design: entry timing, hold horizon, fees.
- Results: Sharpe, Sortino, max DD vs baselines.

8 Robustness and ablations

8.1 Feature pruning & token filtering

- Method: permutation importance per fold $\to$ stable set; drop tokens with heavy imputation.
- Report deltas: pinball, coverage, width.

Sensitivity analyses

Train/cal/test window grid; half-life grid; calibration variant.

9 Discussion

- What the evidence implies for crypto interval forecasting.
- Where each model is preferable; limitations; data quality and missingness.
- Links to literature and practice.

10 Conclusion

- Answers to research questions, contributions, limitations, future work.

11

full token list, missingness thresholds, feature dictionary.

12

full hyperparams per , Optuna spaces, training logs.

13

any additional overlays, reliability/PIT plots, etc.

Meinshausen, N. (2006) ‘Quantile regression forests’, *Journal of Machine Learning Research*, 7, pp. 983–999. Available at: https://doi.org/10.5555/1248547.1248604.