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

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

James Lewis

Table of contents

1 Abstract **4**

2 Introduction **5**
2.0.1 **The Economic Challenge of Forecasting in Emerging Ecosystems** 5
2.0.2 **A Principled Approach: Quantile Regression** 5
2.0.3 **The State of the Art and the Research Gap** 5
2.0.4 **1.4 Problem Formulation: The Solana Ecosystem** 6
2.0.5 **Methodology and Contributions** 6
2.0.6 **Scope and Delimitations** 6
2.0.7 **Research Question** 7
2.0.8 **Dissertation Outline** 7

3 Literature Review **8**
3.0.1 The Challenge: Statistical Properties of Cryptocurrency Returns 8
3.0.2 Approaches to Quantile Estimation 8
3.0.3 Ensuring Rigour: Calibration, Evaluation, and Comparison 10
3.0.4 Methodological Requirements for Robust Time-Series Forecasting 11
3.0.5 Integrating Crypto-Native Data Sources 12
3.0.6 Synthesis and Conclusion 13

4 Data and feature engineering **14**
4.1 Universe and period 14
4.2 Cleaning, alignment, imputation 14
4.3 Feature set 14
4.4 EDA highlights 14

5 Methods **15**
5.1 Rolling evaluation 15
5.2 Models 15
5.2.1 QRF (core) 15
5.2.2 Baselines 15
5.3 Metrics & tests 15
5.4 Key equations 15

6 Results **16**
6.1 Global comparison 16
6.2 Visual overlays (representative tokens) 16

7 Trading application **17**

8 Robustness and ablations **18**
8.1 Feature pruning & token filtering 18

9 Discussion **19**

10 Conclusion **20**

1 Abstract

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

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.

2 Introduction

2.0.1 The Economic Challenge of Forecasting in Emerging Ecosystems

Extreme volatility and non-normal return distributions are defining characteristics of cryptocurrency markets, rendering traditional point forecasts insufficient for robust risk management (Gkillas and Katsiampa, 2018). This challenge is particularly acute for mid-capitalisation tokens within emerging ecosystems like Solana. Unlike large-cap assets, these tokens are subject to rapid narrative shifts and idiosyncratic on-chain dynamics; yet unlike micro-caps, they are liquid enough to attract significant capital. For participants in this space, standard risk models often fail during periods of high network activity or ecosystem-wide events, creating a clear demand for forecasting tools that can dynamically price tail risk.

This dissertation argues that the primary objective must shift from point prediction to **interval forecasting**. By generating calibrated prediction intervals through conditional quantiles, we can capture the asymmetry and tail risk inherent in these volatile assets. A well-calibrated lower quantile serves as a dynamic, forward-looking analogue to Value-at-Risk (VaR) (Engle and Manganelli, 2004), while the upper quantile informs on potential upside, providing a comprehensive basis for sophisticated risk management and tactical decision-making.

2.0.2 A Principled Approach: Quantile Regression

To construct reliable prediction intervals, we adopt the framework of **quantile regression** (Koenker and Bassett, 1978). This approach is formally grounded in the **pinball loss function**, a proper scoring rule that is uniquely minimised when the forecast matches the true conditional quantile of the distribution. For a target outcome y and a forecast for the τ-th quantile, $\hat{q}_\tau$, the loss is:

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

This allows models to be evaluated on two critical properties: **calibration** (does an 80% interval contain the outcome 80% of the time?) and **sharpness** (are the intervals as narrow as possible while maintaining calibration?) (Gneiting and Raftery, 2007). These properties are paramount in crypto markets, where accurately quantifying both risk and opportunity is the basis of effective strategy.

2.0.3 The State of the Art and the Research Gap

The literature on cryptocurrency forecasting has largely bifurcated. One branch applies econometric models like GARCH, which excel at modelling volatility but are constrained by parametric assumptions (Bollerslev, 1986). The other employs machine learning, but predominantly for point forecasting of price or direction (Mcnally, Roche and Caton, 2018). While some studies have applied quantile regression to major crypto-assets (Catania and Sandholdt, 2019), they have typically relied on simpler linear models or have not fully leveraged the rich feature set available from on-chain data. Furthermore, the critical step of post-hoc calibration to ensure nominal coverage guarantees, especially for non-parametric methods, is often overlooked.

This dissertation is designed to address these gaps by making three core contributions: it shifts the focus from point prediction to distributional accuracy; it applies a sophisticated non-parametric methodology to the under-researched domain of mid-cap altcoins; and it integrates a rich, multi-domain feature set that explicitly includes on-chain activity.

2.0.4 1.4 Problem Formulation: The Solana Ecosystem

This study focuses on forecasting **72-hour log-returns** for a universe of mid-cap tokens within the Solana ecosystem, using data aggregated in 12-hour intervals. The asset universe comprises tokens with a market capitalisation exceeding \$30 million, ensuring a focus on liquid yet idiosyncratic assets. The forecasting model is built upon a feature set designed to capture the multi-faceted drivers of returns, spanning five domains: (i) **momentum**, (ii) **volatility**, (iii) **market microstructure**, (iv) **on-chain activity**, and (v) **cross-asset context**.

2.0.5 Methodology and Contributions

The central hypothesis is that the non-linear, interaction-heavy nature of this market demands a non-parametric approach. We propose an adapted **Quantile Regression Forests (QRF)** model (Meinshausen, 2006). QRF was selected as the primary model for several reasons. Firstly, its ensemble nature provides inherent robustness to the noisy predictors common in high-dimensional financial feature sets. Secondly, unlike gradient boosting, QRF's independent tree construction can be less prone to overfitting in non-stationary environments. Finally, its method of estimating quantiles from the full distribution of training samples in terminal nodes is a more direct and empirically stable approach than methods requiring separate models for each quantile. To tailor the model for financial time series, we incorporate several critical enhancements: **time-decay weighting** to prioritise recent data, **volatility regime offsets** to adapt to changing market conditions, and **isotonic regression** to enforce the theoretical non-crossing of quantiles.

We benchmark our adapted QRF against a parametric **Linear Quantile Regression (LQR)** and a powerful **LightGBM** model (Ke *et al.*, 2017) augmented with **conformal prediction** (Romano, Patterson and Candès, 2019). Preliminary results suggest that the primary advantage of the adapted QRF framework lies in its superior ability to synthesise on-chain activity and market microstructure features to anticipate shifts in return distribution skewness—a dynamic that linear models fail to capture.

2.0.6 Scope and Delimitations

This dissertation provides a rigorous methodological and empirical analysis of interval forecasting. It does not aim to develop a complete, production-ready trading system, which would require further considerations such as transaction costs, liquidity constraints, and execution latency. Furthermore, the feature set, while comprehensive, is confined to publicly available market and on-chain data, thereby excluding alternative data sources such as social media sentiment or developer activity metrics, which may also contain predictive information. The findings are specific to the mid-cap tokens within the Solana ecosystem during the observation period and may not be directly generalisable to other blockchains, market-cap tiers, or market regimes without further investigation and potential recalibration.

2.0.7 Research Question

This framework motivates the central research question of this dissertation:

Can an adapted Quantile Regression Forest model deliver sharper and better-calibrated prediction intervals for 72-hour returns of mid-cap Solana tokens compared to standard linear and gradient-boosted quantile regression baselines?

This overarching question is decomposed into four specific, testable hypotheses:

1. Superior Accuracy: The proposed QRF model achieves a lower mean pinball loss across the quantile spectrum than both LQR and LightGBM with conformal prediction.
2. Superior Calibration: The QRF model's empirical coverage rates for 80% and 90% intervals are closer to their nominal levels.
3. Superior Sharpness: The QRF model produces narrower prediction intervals than the conformally-adjusted LightGBM model, without sacrificing calibration.
4. Statistical Significance: The performance improvements offered by QRF are statistically significant as determined by formal tests on pinball loss differentials.

2.0.8 Dissertation Outline

The remainder of this dissertation unfolds as follows. Chapter 2 establishes the theoretical context by reviewing the relevant literature. Chapter 3 details the data pipeline and feature engineering process, while Chapter 4 outlines the core methodology. The empirical analysis begins in Chapter 5 with the main comparative results, which are translated into a practical trading application in Chapter 6 and stress-tested for robustness in Chapter 7. Finally, Chapter 8 discusses the broader implications, leading to the conclusion in Chapter 9, which summarises the contributions and suggests avenues for future research.

3 Literature Review

This chapter provides a critical review of the literature that justifies the methodology of this dissertation. It first establishes the unique statistical properties of cryptocurrency returns that necessitate specialised forecasting approaches. The review then evaluates the relative merits of parametric and non-parametric quantile estimation models, before examining the essential frameworks for robust forecast evaluation, calibration, and comparison. The chapter synthesises these distinct strands of literature to build a coherent argument for the selection of an adapted Quantile Regression Forest as the core model, and for the specific methodological refinements required for its application.

3.0.1 The Challenge: Statistical Properties of Cryptocurrency Returns

The return distributions of cryptocurrencies are characterised by heavy tails, significant skew, and extreme kurtosis relative to traditional assets, reflecting the frequency of large, abrupt price movements (Gkillas and Katsiampa, 2018). This leptokurtosis is compounded by pronounced volatility clustering—periods of relative calm followed by explosive variability—a dynamic exacerbated by the market's continuous operation and fragmented liquidity, which can amplify shocks across uncoordinated venues.

Crucially, this extreme risk is also largely idiosyncratic to the crypto market. Major cryptocurrencies carry substantial tail risk that is not strongly correlated with traditional stock market indices; instead, extreme events are driven by crypto-specific factors such as investor sentiment, regulatory news, or network-level events (Borri, 2019). Furthermore, their returns show little to no exposure to standard macroeconomic risk factors, being influenced instead by internal drivers like network momentum and adoption metrics (Liu and Tsyvinski, 2021). This body of evidence demonstrates that classical financial risk models, with their reliance on Gaussian assumptions and traditional risk factors, are fundamentally misspecified for crypto assets. A credible forecasting framework must therefore abandon these assumptions and be built to incorporate the crypto-native features that drive risk.

3.0.2 Approaches to Quantile Estimation

Given the non-normal character of crypto returns established previously, estimating the full conditional distribution is more informative than forecasting its central tendency. Quantile regression provides a natural framework for this, but the choice between a restrictive parametric model and a flexible non-parametric one is critical.

3.0.2.1 The Parametric Benchmark: Linear Quantile Regression

Quantile regression, introduced by (Koenker and Bassett, 1978), generalises linear regression by estimating conditional quantiles directly. For a given quantile level $\tau \in (0, 1)$, the linear quantile regression (LQR) estimator solves:

$$\hat{\beta}_\tau = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta)$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}\{u < 0\})$ is the **pinball loss function**. While LQR provides a transparent and interpretable benchmark, its fundamental assumption of a fixed linear relationship across all quantiles represents a severe limitation. This rigidity is fundamentally at odds with the non-linear volatility dynamics and abrupt regime shifts that define cryptocurrency markets. Furthermore, the common practical issue of **quantile crossing**—where independently estimated quantile lines intersect—can yield incoherent and unusable interval forecasts unless post-hoc remedies like rearrangement are applied (Chernozhukov, Fernández-Val and Galichon, 2010). These shortcomings do not merely motivate, but necessitate the exploration of more flexible, non-parametric methods.

3.0.2.2 Non-Parametric Solutions: Quantile Regression Forests

As a direct response to the limitations of linear models, Quantile Regression Forests (QRF), proposed by (Meinshausen, 2006), extend the Random Forest algorithm (Breiman, 2001) to estimate the entire conditional distribution. Instead of averaging outcomes in terminal nodes, QRF uses the full empirical distribution of training responses within the leaves to form a predictive distribution, from which conditional quantiles are derived.

This non-parametric approach is inherently well-suited to financial data; it naturally captures complex non-linearities and adapts to heteroskedasticity without pre-specification. However, QRF is not without its own challenges. Its theoretical foundation rests on an assumption of independent and identically distributed (i.i.d.) data—a condition clearly violated by financial time series. A naive application of QRF to time-ordered data can therefore lead to biased estimates. This violation is a central methodological challenge that requires specific adaptations, such as the time-decay weighting and rolling validation schemes discussed later, to apply the model soundly. Furthermore, the accuracy of its tail quantile estimates can degrade if the terminal leaves are sparsely populated, a genuine risk when modelling extreme events. Boosting offers another route to non-parametric quantile estimation, but with contrasting properties.

3.0.2.3 A Boosting Alternative: LightGBM for Quantiles

Gradient boosting presents another powerful non-parametric paradigm. It constructs an ensemble sequentially, with each new tree trained to correct the errors—specifically, the gradients of the loss function—of the preceding models (Friedman, 2001). This methodology can be directly applied to quantile regression by using the pinball loss as the objective. LightGBM (Ke *et al.*, 2017) is a highly efficient and scalable implementation of this idea, making it a formidable baseline.

In sharp contrast to QRF's parallelised construction, boosting's sequential focus on difficult-to-predict instances can yield sharper estimates in the tails. This potential for higher accuracy, however, comes with significant trade-offs. A separate model must typically be trained for each target quantile, imposing a considerable computational burden. The aggressive, error-focused fitting can also produce "ragged" and unstable quantile estimates in regions with sparse data and may overfit transient noise without careful regularisation. Finally, like LQR, independently fitted boosting models are susceptible to the problem of quantile crossing.

3.0.3 Ensuring Rigour: Calibration, Evaluation, and Comparison

Selecting a flexible forecasting model is insufficient on its own; its predictive performance must be evaluated using principled metrics, its outputs calibrated to ensure reliability, and its superiority over alternatives established through formal statistical tests.

3.0.3.1 Proper Scoring and Forecast Evaluation

A principled evaluation of probabilistic forecasts requires the use of **strictly proper scoring rules**, which incentivise the model to report its true belief about the future distribution. For quantile forecasts, the canonical proper scoring rule is the pinball loss (Gneiting and Raftery, 2007). As the metric being directly optimised by the models, it serves as the primary tool for evaluation. However, performance is not a single dimension. The quality of an interval forecast is judged by two distinct and often competing properties: **calibration**, the statistical consistency between the nominal coverage rate (e.g., 90%) and the empirical frequency of outcomes falling within the interval; and **sharpness**, the narrowness of the interval. An ideal forecast is one that is maximally sharp, subject to being well-calibrated. However, a model optimised on a proper score is not inherently guaranteed to be well-calibrated in finite samples. This gap between theoretical optimisation and empirical reliability motivates the use of formal calibration techniques.

3.0.3.2 Achieving "Honest" Intervals: Conformal Prediction

Conformal prediction provides a distribution-free framework to correct for such miscalibration. Specifically, Conformalized Quantile Regression (CQR) (Romano, Patterson and Candès, 2019) provides a mechanism to adjust a base model's quantile forecasts to achieve guaranteed marginal coverage. It uses a hold-out calibration set to compute a conformity score based on model errors, which is then used to adjust the width of future prediction intervals. While the underlying exchangeability assumption is violated in time series, employing a rolling calibration window of recent data provides a practical and widely used compromise to adapt the procedure to non-stationary environments.

3.0.3.3 Statistically Significant Comparisons: The Diebold-Mariano Test

To move beyond descriptive comparisons of average loss, formal statistical tests are required to determine if the performance difference between two models is significant. The **Diebold-Mariano (DM) test** (Diebold and Mariano, 1995) provides a standard framework for this, assessing the null hypothesis of equal predictive accuracy. The test statistic is given by:

$$DM = \frac{\bar{d}}{\sqrt{\hat{\mathrm{Var}}(\bar{d})}}$$

where $\bar{d}$ is the mean loss differential between two models. For the multi-step, overlapping forecasts used in this project, the sequence of loss differentials will be autocorrelated by construction. It is therefore critical to use a heteroskedasticity and autocorrelation consistent (HAC) variance estimator, as recommended by (West, 1996), to ensure valid statistical inference.

3.0.4 Methodological Requirements for Robust Time-Series Forecasting

The foundational literature establishes the potential of non-parametric models, but their successful application to volatile, non-stationary financial time series is contingent upon a number of specific methodological adaptations. This section reviews the literature concerning these essential requirements, from ensuring the logical coherence of predictions to adapting models to the temporal dynamics of the data.

3.0.4.1 Ensuring Coherent Predictions: Non-Crossing Quantiles

Models that estimate quantiles independently, such as LQR and standard gradient boosting, are susceptible to the critical failure of **quantile crossing**. This occurs when, for instance, a predicted 90th percentile falls below the predicted 50th percentile, yielding a logically incoherent and unusable conditional distribution. To rectify this, (Chernozhukov, Fernández-Val and Galichon, 2010) proposed a post-processing "rearrangement" technique. This method applies isotonic regression to the initially estimated quantile function, projecting the unconstrained predictions onto the space of valid, non-decreasing distribution functions. This ensures the monotonicity of the quantile curve and is a critical step for producing valid prediction intervals.

3.0.4.2 Adapting to Non-Stationarity and Temporal Dependence

Financial time series are fundamentally non-stationary and autocorrelated, violating the i.i.d. assumption that underpins many machine learning models. Two distinct but related adaptations are required to address this.

First, to handle **non-stationarity** such as volatility clustering, the model must prioritise more recent information. The literature supports the use of **time-decay sample weights** to achieve this. (Taylor, 2008), for example, introduced exponentially weighted quantile regression for Value-at-Risk estimation, demonstrating that up-weighting recent observations yields more responsive and accurate tail forecasts in changing market conditions.

Second, to handle **temporal dependence**, model evaluation and hyperparameter tuning must respect the chronological order of the data. Standard k-fold cross-validation is invalid for time series, as it can lead to lookahead bias and produce misleadingly optimistic performance estimates. The literature therefore strongly advocates for rolling-origin or blocked cross-validation schemes, which preserve the temporal sequence by training only on past data to forecast the future, thereby simulating a live forecasting environment (Bergmeir, Hyndman and Koo, 2018).

3.0.4.3 Correcting for Bias and Ensuring Empirical Calibration

Even correctly specified quantile models can exhibit systematic biases in finite samples. As (Bai *et al.*, 2021) have shown, linear quantile regression can suffer from a theoretical **undercoverage bias**, where a nominal 90% interval may contain the true outcome significantly less than 90% of the time due to estimation error. This problem motivates the necessity of post-hoc calibration.

While the CQR framework discussed previously is one such solution, the literature offers several alternatives. Methods like the Jackknife+ (Barber *et al.*, 2021) and residual bootstraps provide different mechanisms for constructing prediction intervals with more reliable coverage properties. The existence of this rich literature on calibration highlights a crucial principle for risk management applications: a model's raw output cannot be taken at face value. An explicit calibration step is required to correct for inherent biases and ensure the resulting prediction intervals are empirically "honest".

3.0.5 Integrating Crypto-Native Data Sources

The literature on cryptocurrency risk factors makes it clear that models confined to historical price data are insufficient. The unique nature of blockchain-based assets provides a rich set of crypto-native data sources that are essential for capturing the specific drivers of risk and return in this asset class.

3.0.5.1 Market Microstructure and Liquidity

Like traditional markets, cryptocurrency price dynamics are influenced by liquidity and trading frictions. Empirical studies have documented that periods of market stress coincide with widening bid-ask spreads and evaporating order book depth (Dyhrberg, 2016). Furthermore, the on-chain nature of transactions introduces unique microstructural features, such as network congestion and transaction fees, which can impact market liquidity and price formation (Easley, O'Hara and Basu, 2019). Incorporating proxies for these effects is crucial, as it allows a model to dynamically adjust its estimate of uncertainty; for instance, by widening its prediction intervals in response to deteriorating market liquidity, thereby anticipating volatility spikes.

3.0.5.2 On-Chain Activity and Network Fundamentals

Blockchains provide a transparent ledger of network activity, offering powerful proxies for an asset's fundamental adoption and utility. Metrics such as the growth in active addresses, on-chain transaction counts, and, in the context of decentralised finance (DeFi), the Total Value Locked (TVL) in smart contracts, can signal shifts in investor sentiment and capital flows. Empirical studies consistently find that models augmented with on-chain metrics significantly outperform those based only on historical prices, as this data provides unique information about network health and demand (Sebastião and Godinho, 2021). These features allow a model to condition its forecasts on the fundamental state of the network, potentially informing not just the location but also the shape of the predictive distribution.

3.0.5.3 Cross-Asset Spillovers and Systemic Risk

The cryptocurrency market is a highly interconnected system where shocks to major assets like Bitcoin and Ethereum often propagate to smaller altcoins. This "connectedness" has been formally measured, showing significant return and, particularly, volatility spillovers from market leaders to the rest of the ecosystem (Diebold and Yilmaz, 2014; Koutmos, 2018). This implies that the risk of an individual token is not purely idiosyncratic; it is also a function of the broader crypto market's state. Consequently, any forecasting model that treats a token in isolation is fundamentally misspecified and is likely to underestimate systemic risk. A robust framework must therefore account for these cross-asset influences.

3.0.6 Synthesis and Conclusion

This review has established a clear and compelling justification for the methodology adopted in this dissertation. The unique statistical properties of cryptocurrency returns—heavy tails, volatility clustering, and dependence on idiosyncratic, on-chain factors—render traditional parametric models inadequate. This failure necessitates the use of flexible, non-parametric methods, for which Quantile Regression Forests are a logical choice, given their ability to capture complex, non-linear relationships without restrictive distributional assumptions.

However, the literature also makes it clear that a naive application of any such model would be insufficient. A credible forecasting framework requires a series of specific, evidence-based adaptations. The need to adapt to non-stationarity justifies the use of time-decay weighting. The imperative for valid, coherent predictions necessitates post-processing to enforce non-crossing quantiles. The requirement for reliable out-of-sample evaluation mandates the use of rolling cross-validation. Finally, the well-documented tendency for quantile models to mis-calibrate compels the integration of a formal calibration step to ensure the final prediction intervals are empirically valid.

By synthesising these distinct strands of literature—from model selection to time-series adaptation and calibration—this project constructs an integrated and methodologically robust framework. This framework is specifically designed to address the multifaceted challenges of interval forecasting in the volatile and rapidly evolving cryptocurrency market.

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 → 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.

Bai, Y. *et al.* (2021) ‘Understanding the under-coverage bias in uncertainty estimation’, *arXiv preprint arXiv:2104.05818* [Preprint].

Barber, R.F. *et al.* (2021) ‘Predictive inference with the jackknife+’, *The Annals of Statistics*, 49(1), pp. 486–507.

Bergmeir, C., Hyndman, R.J. and Koo, B. (2018) ‘A note on the validity of cross-validation for evaluating autoregressive time series prediction’, *Computational Statistics & Data Analysis*, 120, pp. 70–85.

Bollerslev, T. (1986) ‘Generalized autoregressive conditional heteroskedasticity’, *Journal of Econometrics*, 31(3), pp. 307–327.

Borri, N. (2019) ‘Conditional tail-risk in cryptocurrency markets’, *Journal of Empirical Finance*, 50, pp. 1–19.

Breiman, L. (2001) ‘Random forests’, *Machine Learning*, 45(1), pp. 5–32.

Catania, L. and Sandholdt, M. (2019) ‘Bitcoin at high frequency’, *Journal of Risk and Financial Management*, 12(1), p. 36.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010) ‘Quantile and probability curves without crossing’, *Econometrica*, 78(3), pp. 1093–1125.

Diebold, F.X. and Mariano, R.S. (1995) ‘Comparing predictive accuracy’, *Journal of Business & Economic Statistics*, 13(3), pp. 253–263.

Diebold, F.X. and Yilmaz, K. (2014) ‘On the network topology of variance decompositions: Measuring the connectedness of financial firms’, *Journal of Econometrics*, 182(1), pp. 119–134.

Dyhrberg, A.H. (2016) ‘Bitcoin, gold and the dollar–a GARCH volatility analysis’, *Finance Research Letters*, 16, pp. 85–92.

Easley, D., O’Hara, M. and Basu, S. (2019) ‘From mining to markets: The evolution of bitcoin transaction fees’, *Journal of Financial Economics*, 134(1), pp. 91–109.

Engle, R.F. and Manganelli, S. (2004) ‘CAViaR: Conditional autoregressive value at risk by regression quantiles’, *Journal of Business & Economic Statistics*, 22(4), pp. 367–381.

Friedman, J.H. (2001) ‘Greedy function approximation: A gradient boosting machine’, *The Annals of Statistics*, 29(5), pp. 1189–1232.

Gkillas, K. and Katsiampa, P. (2018) ‘An application of extreme value theory to cryptocurrencies’, *Economics Letters*, 164, pp. 109–111.

Gneiting, T. and Raftery, A.E. (2007) ‘Strictly proper scoring rules, prediction, and estimation’, *Journal of the American Statistical Association*, 102(477), pp. 359–378.

Ke, G. *et al.* (2017) ‘LightGBM: A highly efficient gradient boosting decision tree’, in *Advances in neural information processing systems 30.* Curran Associates, Inc., pp. 3146–3154.

Koenker, R. and Bassett, G. (1978) ‘Regression quantiles’, *Econometrica*, 46(1), pp. 33–50.

Koutmos, D. (2018) ‘Return and volatility spillovers among cryptocurrencies’, *Economics Letters*, 173, pp. 122–127.

Liu, Y. and Tsyvinski, A. (2021) ‘Risks and returns of cryptocurrency’, *The Review of Financial Studies*, 34(6), pp. 2689–2727.

Mcnally, S., Roche, J. and Caton, S. (2018) ‘Predicting the price of bitcoin using machine learning’, in *2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP).* IEEE, pp. 339–343.

Meinshausen, N. (2006) ‘Quantile regression forests’, *Journal of Machine Learning Research*, 7, pp. 983–999.

Romano, Y., Patterson, E. and Candès, E.J. (2019) ‘Conformalized quantile regression’, in *Advances in neural information processing systems 32.* Curran Associates, Inc., pp. 3543–3553.

Sebastião, H. and Godinho, P. (2021) ‘Forecasting cryptocurrency prices with blockchain network metrics’, *Expert Systems with Applications*, 177, p. 114944.

Taylor, J.W. (2008) ‘Using exponentially weighted quantile regression to estimate value at risk and expected shortfall’, *Journal of Financial Econometrics*, 6(3), pp. 382–406.

West, K.D. (1996) ‘Asymptotic inference about predictive ability’, *Econometrica*, 64(5), pp. 1067–1084.