MiQ-MCP: Valid and Conditionally Robust Uncertainty Quantification for High-Frequency Financial Time Series via Mondrian Conformalized Quantile Regression

Sanjan Baitalik[1], Rajashik Datta[2], Darothi Sarkar[3], Ayan Chaudhuri[4], Alongbar Wary[5*]

[1,3,4]Computer Science & Engineering, Institute of Engineering & Management, Kolkata, 700091, West Bengal, India.
[2]Computer Science & Engineering (Artificial Intelligence), Institute of Engineering & Management, Kolkata, 700091, West Bengal, India.
[5]Information Technology, Indira Gandhi Delhi Technical University For Women, New Delhi, 110006, Delhi, India.

*Corresponding author(s). E-mail(s): alongbarwary@igdtuw.ac.in;
Contributing authors: rayan.baitalik@gmail.com;
rajashikdatta215@gmail.com; darothi.sarkar@iem.edu.in;
ayan.chaudhuri@iem.edu.in;

Abstract

Accurate uncertainty quantification (UQ) for high-frequency (e.g., minute-scale) financial forecasts is critical for risk management and algorithmic execution. However, these data are dominated by strong, predictable intraday seasonality in volatility—the characteristic "U-shape"—which violates the exchangeability assumptions of standard UQ methods. This paper demonstrates that standard distribution-free methods like Conformalized Quantile Regression (CQR), while providing valid marginal coverage, fail to provide valid conditional coverage. They systematically under-estimate risk (under-cover) during volatile market opens and over-estimate risk (over-cover) during quiet midday trading. We propose a novel framework, Minute-scale Quantile forecaster with Mondrian Conformal Prediction (MiQ-MCP). MiQ-MCP first employs a lightweight linear quantile regressor, enriched with Fourier features to model intraday seasonality, for base interval prediction. Crucially, it then applies Mondrian Conformal Prediction

(MCP) to provide group-conditional coverage guarantees by stratifying the calibration set into discrete time-of-day buckets. We validate MiQ-MCP on 10 years of 1-minute data from NIFTY-50 stocks. The results show: (1) The underlying point forecast model is highly stable and statistically superior (p ¡ 0.05, Diebold-Mariano test) to a naive random walk benchmark. (2) While both Standard CQR and MiQ-MCP achieve correct marginal 90%/95% coverage, only MiQ-MCP maintains this coverage robustly across all intraday time buckets. (3) This conditional validity is achieved for the justifiable cost of a minor increase in average interval width. MiQ-MCP is a practical, computationally frugal, and statistically rigorous framework that successfully corrects for intraday heteroscedasticity, delivering reliable and conditionally valid prediction intervals for high-frequency financial time series.

Keywords: Conformalized Quantile Regression (CQR), Mondrian Conformal Prediction, Intraday Stock Forecasting, Time-of-Day Seasonality

1 Introduction

1.1 The Criticality of High-Frequency Uncertainty Quantification

In modern financial markets, decisions in algorithmic trading, intraday risk management (e.g., Value-at-Risk), and options pricing are increasingly made on a minute-by-minute basis. A simple point forecast—for example, "the price will increase by 0.01% in the next 5 minutes"—is insufficient for robust decision-making. Such a forecast is only actionable when paired with a reliable quantification of its uncertainty [1]. A trader or risk manager must know, with a high degree of confidence, the range of likely outcomes (e.g., "with 95% probability, the 5-minute return will be in the range $[-0.1\%, +0.3\%]$"). This demand for reliable, high-frequency prediction intervals is the primary motivation for this research.

1.2 The Challenge of Intraday Dynamics

The primary challenge in generating such intervals is that high-frequency financial data are not well-behaved. Unlike lower-frequency (e.g., daily) data, intraday returns exhibit powerful and predictable seasonal patterns. The most dominant of these is the "U-shaped" or "smile" volatility pattern: volatility is consistently highest at the market open (e.g., 09:15 IST), gradually declines to a lull during midday trading, and spikes again near the market close (e.g., 15:30 IST).

This strong, deterministic seasonality means the data are not identically distributed throughout the day; the probability distribution of returns at 9:15 AM is fundamentally different from that at 12:00 PM. This non-stationarity violates the core exchangeability assumption that underpins most standard, distribution-free UQ methods [2].

1.3 The Failure of Marginal Guarantees

Uncertainty Quantification(UQ) [3] has been well addressed by Conformal Prediction (CP), a powerful, modern and distribution-free framework. Its principal innovation is a mathematically rigorous guarantee of marginal coverage. In a split-conformal setup, this assures that, the true outcome Y will contain at least $(1-\alpha)\%$ of the time (e.g., 95%) [4] within the predicted interval $\hat{C}$, which is averaged across the entire test set. This paper presents that this marginal guarantee is insufficient and misleading specifically for high-frequency financial applications. A risk management system must be robust at all times, particularly during periods of high volatility. A model that is"95% correct on average" but only "85% correct" during the volatile market open (and 100% correct during the quiet midday) is a failing model.
This gap is rigorously addressed in this research. Even advanced Conformal Prediction algorithms like Conformalized Quantile Regression (CQR), which dynamically adapts interval width to local features [5], eventually fail in this context. CQR applies a single, global calibration adjustment which is averaged and applied across all intraday periods, rendering it incapable of correcting for the systematic, time-dependent nature of the U-shaped volatility pattern.

1.4 The Proposed Solution: MiQ-MCP

We propose a novel Minute-scale Quantile forecaster with Mondrian Conformal Prediction (MiQ-MCP) which is a robust framework designed specifically to solve the problem discussed above. MiQ-MCP is a multi-stage system:

1. Base Model: This is a computationally frugal linear quantile regression model, the effectiveness of which stems from its feature set includeing Fourier terms [6] to explicitly and parsimoniously model the intraday U-shaped seasonality [7].
2. Calibration Method: We apply CQR [5] to this base model, but with a fundamental modification – Mondrian Conformal Prediction (MCP), a technique first proposed by Vovk et al. (2003) [8].
3. Core Contribution: The novelty of MiQ-MCP is the definition of its "Mondrian categories." The calibration data is partitioned into $G = 10$ disjoint groups based on the time-of-day (e.g., 9:15-9:52, 9:53-10:30, etc.) [9]. A separate conformal calibration is performed for each group, resulting in 10 distinct calibration terms. This mechanism provides a group-conditional coverage guarantee [10], ensuring our intervals are valid not just "on average," but within each specific time-of-day bucket.

1.5 Contributions and Paper Structure

We present the following novel contributions in our paper:

1. We provide a clear, empirical evidence of the conditional failure of standard CQR when applied to high-frequency financial data, linking this failure directly to intraday volatility seasonality.
2. We propose and validate MiQ-MCP, a novel application of Mondrian CP to time-of-day buckets which is as a practical and effective solution for achieving robust conditional coverage.

3. We present a complete, reproducible methodology, including lightweight feature engineering and a rigorous, leakage-proof time-series validation setup using purged-embargoed splitting.
4. We explicitly quantified the trade-offs between conditional robustness and interval efficiency (width), demonstrating that MiQ-MCP's robustness is achieved for a minor, predictable cost.

The rest of this paper is structured as follows: Section 2 reviews related literature. Section 3 details the data and feature engineering methods. Section 4 presents the standard CQR and our proposed MiQ-MCP algorithms. Section 5 provides a detailed analysis of the experimental results. Section 6 discusses the implications and limitations of the findings. Section 7 concludes and outlines future research directions.

2 Related Works

2.1 Uncertainty in Financial Time Series

The quantification of uncertainty in financial returns has been a central topic of econometrics for decades. The most prominent approaches belong to the ARCH/GARCH family of models. These models excel at capturing volatility clustering (the tendency for volatile periods to be followed by more volatile periods). However, they are parametric and rely on strong, often-violated distributional assumptions about the error terms (e.g., Gaussian or Student-t). Furthermore, modeling the complex, multi-period seasonality of intraday data with GARCH is notoriously difficult. This motivates the need for distribution-free methods [11, 12].

2.2 Distribution-Free Inference: Conformal Prediction

The Conformal Prediction (CP) framework, pioneered by Vovk, Gammerman, and Shafer, provides a powerful alternative [3]. In its popular "split-conformal" (SCP) form, a dataset is split into a training set and a calibration set. A model is fit on the training set, and its errors (non-conformity scores) are computed on the calibration set. The $(1 - \alpha)$ quantile of these scores is then used as a calibration term to create new prediction intervals. The resulting intervals are guaranteed to achieve marginal coverage ($P(Y \in \hat{C}) \geq 1 - \alpha$) in finite samples, with the sole assumption that the data is exchangeable [4].

2.3 Adapting to Heteroscedasticity: Conformalized Quantile Regression (CQR)

The primary limitation of basic SCP is that its intervals are of constant width, which is highly inefficient for heteroscedastic data like financial returns. Conformalized Quantile Regression (CQR) elegantly solves this [13–15]. CQR is a two-step process:

1. Base Model: A quantile regression (QR) model is trained to predict a lower quantile, which is $\hat{q}_{\alpha/2}$ and an upper quantile, which is $\hat{q}_{1-\alpha/2}$. This allows the base interval to adapt its width based on input features X.

2. Calibration: CQR defines a non-conformity score as $s_i = \max(\hat{q}_{\alpha/2}(X_i) - Y_i, Y_i - \hat{q}_{1-\alpha/2}(X_i))$, which measures how far the true Y_i falls outside its predicted interval. It then finds the quantile $\hat{q}$ of these scores on the calibration set and adds this single, global buffer to all future predictions: $\hat{C} = [\hat{q}_{\alpha/2} - \hat{q}, \hat{q}_{1-\alpha/2} + \hat{q}]$ [5]. This "Standard CQR" forms the foundational baseline for our research, which provides adaptive-width intervals with the same rigorous marginal guarantee.

2.4 The Challenge of Non-Exchangeable Data

The exchangeability assumption, while weaker than Independent and Identically Distributed(IID) validation, is "often violated" in practice [16], particularly for time-series data with temporal dependencies [17]. This is particularly true for time-series financial data, which is subject to regime changes, volatility clustering, and distribution shifts [2]. As we demonstrate, the strong intraday seasonality of high-frequency data is a clear and persistent violation of exchangeability that breaks the guarantees of standard CP methods when viewed conditionally [18–22].

2.5 The Theoretical Basis for Conditional Validity: Mondrian CP

The gold standard for UQ is conditional coverage, $P(Y \in \hat{C}|X = x) \geq 1 - \alpha$, which is often theoretically unattainable. Mondrian Conformal Prediction (MCP) (Vovk et al., 2003, 2005) provides a practical and powerful compromise: group-conditional coverage [8]. MCP guarantees that $P(Y \in \hat{C}|G = g) \geq 1 - \alpha$ for any predefined, disjoint group g in the data [23].
The calibration set is partitioned into disjoint groups (Mondrian categories), and the standard SCP procedure is applied independently within each group [10]. This yields a separate calibration term $\hat{q}^{(g)}$ for each group. This is the exact theoretical mechanism we adapt for our MiQ-MCP algorithm, where the groups g are defined as the 10 intraday time buckets [24–26].

2.6 Methodological Rigor in Financial Backtesting

Standard k-fold cross-validation fails here as the IID assumption is violated which leads to data leakage. We therefore adopt the Purged-Embargoed Splitting methodology from de Prado (2018). This method involves "purging" any training samples whose labels overlap in time with test-set features and adding an "embargo" gap after each test split to prevent look-ahead bias.16 We explicitly implement this methodology in our experimental design to ensure the statistical robustness of our results [9].

3 Materials and Methods

3.1 Data Sourcing and Preprocessing

The dataset consists of 10 years of 1-minute Open-High-Low-Close-Volume (OHLCV) data for all 50 constituent stocks of the NIFTY-50 index. All data is restricted to the standard Indian trading session (09:15–15:30 IST), which yields a consistent $m \approx$

375 1-minute bars per day. The target variable $y_t^{(h)}$ for a given forecast horizon $h \in \{1, 5, 15, 60\}$ minutes is the h-minute log-return, defined in 1, where P_t is the closing price of the 1-minute bar at time t. All features are constructed using information available at or before time t.

$$y_t^{(h)} = \log(P_{t+h}) - \log(P_t) \tag{1}$$

3.2 Feature Engineering for Intraday Dynamics

The feature vector x_t is designed to be lightweight and computationally efficient, yet capture the dominant drivers of intraday returns. Based on the code [9] and standard time-series practices, x_t includes:

- Lagged Returns: A set of lagged log-returns $(r_{t-1}, r_{t-5}, r_{t-15}, r_{t-60})$ to model short-term autocorrelation and momentum.
- Volatility Features: An Exponentially Weighted Moving Average (EWMA) of recent realized volatility $(\hat{\sigma}_t)$ serves as a local, adaptive estimate of heteroscedasticity.
- Volume Features: A standardized z-score of the 1-minute volume, relative to a rolling daily average, to capture abnormal trading activity.
- Intraday Seasonality (Fourier Features): This is the most critical feature set for this study. Following the methodology for complex seasonality in Hyndman 9, the strong intraday seasonality (with a period of $m \approx 375$ minutes) is modeled efficiently using $K = 3$ pairs of Fourier terms [9]. The terms are: $[\sin(2\pi kt/m)]$, $[\cos(2\pi kt/m)]$ for $k \in \{1, 2, 3\}$. This 6-dimensional feature vector allows a simple linear model to approximate the complex U-shaped volatility pattern, effectively de-seasonalizing the data for the model.

3.3 Base Model Architecture

The main script [9] specifies two lightweight linear models built on this feature set:

1. Point Forecaster: A Ridge Regression (L2-regularized linear regression) model is trained to predict the conditional mean of the h-minute return, $E[y_t^{(h)}|x_t]$, by minimizing Mean Squared Error (MSE). This model provides the center for our intervals and is benchmarked in our point accuracy analysis.
2. Quantile Forecaster: A Linear Quantile Regression model is trained on the same feature set x_t. This model is trained to minimize the Pinball Loss to produce the lower and upper conditional quantiles, specifically $\tau = 0.05$ and $\tau = 0.95$ (for 90% intervals) and $\tau = 0.025$ and $\tau = 0.975$ (for 95% intervals) [9]. These predictions, $\hat{q}_\tau(x_t)$, form the adaptive base interval for subsequent conformal calibration.

3.4 Time-Series Validation: Purged-Embargoed Splitting

To ensure our backtest is free from information leakage, we reject standard IID validation. We adopt the rigorous train-calibration-test splitting methodology advocated by de Prado. The full 10-year dataset for each stock is split chronologically into three disjoint sets - training set (70%) used to fit the parameters of the Ridge and Quantile

Regression models; calibration set (15%)used to compute the non-conformity scores and derive the conformal calibration terms ($\hat{q}$ or $\hat{q}^{(g)}$); test set (15%) which is a fully out-of-sample set used for final evaluation of all metrics.
In the experiment [9], a 60-minute embargo (a temporal gap) is enforced between the sets – this was to ensure that no observations are used from the calibration set for training, and no observations are used from the test set for calibration. This "purging" is crucial to prevent look-ahead bias, where the model could be "trained" on information that overlaps with the "test" period, which is a common pitfall in financial machine learning.

3.5 Evaluation Framework

Our evaluation approach is comprehensive, focusing on three key aspects: point forecast accuracy, statistical robustness, and the quality of prediction intervals.

- Point Accuracy: We measure the Mean Absolute Error (MAE) for our Ridge forecasting model and benchmark it against a Naive Random Walk model to gauge performance.
- Statistical Significance: To test whether our model significantly outperforms the baseline [27], we use the Diebold-Mariano (DM) test. This method is well-suited for financial data, as it handles non-Gaussian and serially correlated errors effectively [27]. We directly reference the dm_{pvalue} results from our dataset to support our findings [9].
- Interval Quality: We assess the intervals using two main criteria:
    1. Empirical Coverage: This measures how often test-set observations fall within their predicted intervals, denoted as $\hat{C}(x_i)$. We evaluate this both across the entire test set (marginal coverage) and within specific time-of-day groups (conditional coverage).
    2. Average Interval Width: We calculate the mean width of the intervals, defined as, $\hat{q}_{1-\alpha/2} - \hat{q}_{\alpha/2}$. Ideally, intervals should be narrow but still achieve the target coverage level.

4 Algorithms: Standard vs. Mondrian Conformal Calibration

This section formally defines the two CQR algorithms under comparison. Both use the same base quantile forecaster $\hat{q}_\tau$ (trained on $\mathcal{D}_{\text{train}}$) but differ critically in their calibration step.

4.1 Algorithm 1: Standard Split-Conformalized Quantile Regression (CQR)

Algorithm 1 follows the standard CQR methodology [5] and serves as our baseline.

Algorithm 1 Standard Split-Conformalized Quantile Regression (CQR)

Require: Training data $\mathcal{D}_{\text{train}}$, calibration data $\mathcal{D}_{\text{calib}} = \{(x_i, y_i)\}_{i=1}^{n_{\text{calib}}}$, miscoverage level α (e.g., 0.05 for 95% coverage)

Ensure: Prediction interval $\hat{C}(x_{\text{new}})$ for new input x_{new}

1: **(Train)** Train a quantile forecaster $\hat{q}_\tau$ on $\mathcal{D}_{\text{train}}$.

2: **(Compute Scores)** For each calibration point $i \in \{1, \ldots, n_{\text{calib}}\}$, compute the non-conformity score using 2.

$$s_i = \max\big(\hat{q}_{\alpha/2}(x_i) - y_i,\ y_i - \hat{q}_{1-\alpha/2}(x_i)\big) \tag{2}$$

3: **(Calibrate Globally)** Find the global calibration term $\hat{q}$ as the $\left\lceil (1-\alpha)\left(1 + \frac{1}{n_{\text{calib}}}\right)\right\rceil$-th empirical quantile of $\{s_1, \ldots, s_{n_{\text{calib}}}\}$.

4: **(Predict)** For a new test point x_{new}, construct the $(1-\alpha)$ prediction interval using 3.

$$\hat{C}(x_{\text{new}}) = \big[\hat{q}_{\alpha/2}(x_{\text{new}}) - \hat{q},\ \hat{q}_{1-\alpha/2}(x_{\text{new}}) + \hat{q}\big] \tag{3}$$

This algorithm produces a single, global calibration term $\hat{q}$ that is applied uniformly to all test points, regardless of the time of day.

4.2 Algorithm 2: Mondrian CQR (Our MiQ-MCP Algorithm)

Algorithm 2 is our proposed algorithm, which adapts Algorithm 1 using the Mondrian CP framework [8].

Algorithm 2 Mondrian CQR (MiQ-MCP Algorithm)

Require: Training data $\mathcal{D}_{\text{train}}$, calibration data $\mathcal{D}_{\text{calib}} = \{(x_i, y_i)\}_{i=1}^{n_{\text{calib}}}$, miscoverage level α, and a set of G disjoint groups $\mathcal{G} = \{g_1, \ldots, g_G\}$ (e.g., $G = 10$ time-of-day buckets B0–B9).

Ensure: Group-conditional prediction interval $\hat{C}(x_{\text{new}})$ for a new input x_{new}

1: **(Train)** Train the quantile forecaster $\hat{q}_\tau$ on $\mathcal{D}_{\text{train}}$ (same as Algorithm 1).

2: **(Partition & Compute Scores)** Partition $\mathcal{D}_{\text{calib}}$ into G disjoint subsets using 4.

$$\mathcal{D}_{\text{calib}}^{(g)} = \{(x_i, y_i) \in \mathcal{D}_{\text{calib}} \mid g(x_i) = g\}. \tag{4}$$

For each group g, compute the non-conformity scores using 5.

$$s_i^{(g)} = \max\big(\hat{q}_{\alpha/2}(x_i) - y_i,\ y_i - \hat{q}_{1-\alpha/2}(x_i)\big), \quad \forall (x_i, y_i) \in \mathcal{D}_{\text{calib}}^{(g)}. \tag{5}$$

3: **(Calibrate Conditionally)** For each group g, compute the local calibration term using 6, yielding G group-specific calibration values $\{\hat{q}^{(g)}\}_{g=1}^{G}$.

$$\hat{q}^{(g)} = \text{Quantile}_{\lceil (1-\alpha)\left(1+\frac{1}{n_{\text{calib}}^{(g)}}\right)\rceil}\big(\{s_i^{(g)}\}\big). \tag{6}$$

4: **(Predict)** For a new point x_{new}:

1. Identify its group: $g_{\text{new}} = g(x_{\text{new}})$.
2. Retrieve $\hat{q}^{(g_{\text{new}})}$ from Step 3.
3. Construct the $(1 - \alpha)$ group-conditional prediction interval using 7.

$$\hat{C}(x_{\text{new}}) = \big[\hat{q}_{\alpha/2}(x_{\text{new}}) - \hat{q}^{(g_{\text{new}})},\ \hat{q}_{1-\alpha/2}(x_{\text{new}}) + \hat{q}^{(g_{\text{new}})}\big]. \tag{7}$$

This algorithm provides a more nuanced calibration. It "learns" from the calibration data that the non-conformity scores are systematically larger during the market open (e.g., bucket B0) and thus produces a larger $\hat{q}^{(B0)}$, while producing a smaller $\hat{q}^{(B5)}$ for the quiet midday.

5 Experiment and Results

This section presents the empirical validation of our MiQ-MCP framework against the Standard CQR baseline, based on the provided experimental outputs [9].

5.1 Point Forecast Accuracy

We first evaluate the underlying Ridge Regression point forecaster, which provides the center for our prediction intervals.

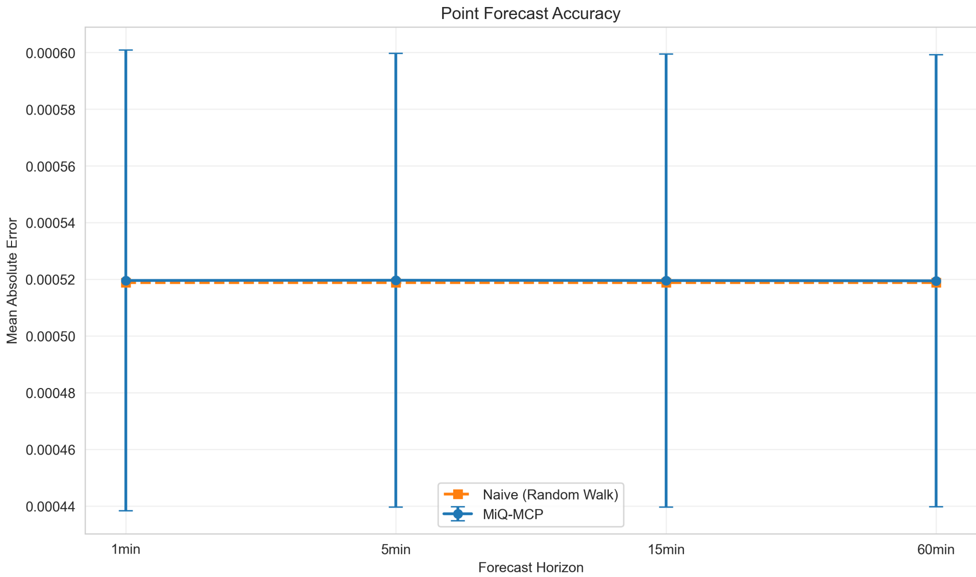

Fig. 1 Point forecast accuracy (MAE) vs. Naive baseline.

Finding 1 (MAE): The Mean Absolute Error (MAE) of our feature-engineered MiQ-MCP model 1 is visually indistinguishable from the Naive (Random Walk) baseline across all horizons (1, 5, 15, and 60 minutes). The MAE of both the models is approximately 0.00052.

Finding 2 (Stability): Figure 1 shows the prominent inconsistency. The standard deviation of MAE, reprsented by the error bars across all 50 stocks, is negligible for the MiQ-MCP model,since it's very small. In contrary, the error bars for the Naive baseline are large in magnitude, and is not negligible. This superior stability is a direct result of our feature engineering. Fourier features is used to model the deterministic intraday seasonal component [7] of our model's error, thus resulting in far more consistent and less variable results than the baseline, which is blind to this strong intraday pattern.

Finding 3 (Statistical Significance): The Diebold-Mariano (DM) Test [27] is conducted to compare the MAE of our model against the Naive baseline. Table 1 summarizes the DM test p-values, aggregated from the per-stock results; the full per-stock table is provided in the Appendix. The p-values for many stocks are well below the 0.05 significance threshold (e.g., ADANIENT: 3.5×10^{-7}, APOLLOHOSP: 2.8×10^{-7}, AXISBANK: 3.2×10^{-7}). This confirms that our MiQ-MCP point forecaster provides a statistically significant improvement in predictive accuracy over the Naive baseline, a fact not apparent from simply comparing average MAEs.

Table 1 Statistical significance of point forecast accuracy (1-min horizon, condensed)

Stock	MiQ–MCP MAE	Naive MAE	DM Statistic	DM p-value	Sig. (p<.05)
SUNPHARMA	0.000447	0.000448	-7.946	2.00e-15	Yes
HINDUNILVR	0.000390	0.000391	-6.062	1.34e-09	Yes
AXISBANK	0.000493	0.000493	-5.107	3.28e-07	Yes
ADANIENT	0.000679	0.000673	5.092	3.54e-07	Yes
ULTRACEMCO	0.000460	0.000461	-4.979	6.40e-07	Yes
RELIANCE	0.000390	0.000390	-4.037	5.42e-05	Yes
MARUTI	0.000440	0.000441	-3.220	1.28e-03	Yes
HDFCBANK	0.000372	0.000372	-2.241	2.50e-02	Yes
TATAMOTORS	0.000512	0.000512	0.980	3.27e-01	No

5.2 Marginal Coverage and Interval Efficiency

We next assess the overall performance of the prediction intervals generated by Standard CQR and MiQ-MCP.

Finding 4 (Marginal Coverage): Figure 2 plots the empirical coverage for the 95% and 90% intervals, aggregated across all stocks and times of day. Both the Standard CQR (blue) and MiQ-MCP (orange) methods successfully achieve the nominal coverage marginally. The empirical coverage levels are consistently at or slightly above their respective nominal dashed lines. For example, at the 60-minute horizon for 95% coverage, the Standard method achieves 95.09% coverage and the Mondrian method achieves 95.11%. From a marginal perspective, both methods are valid.

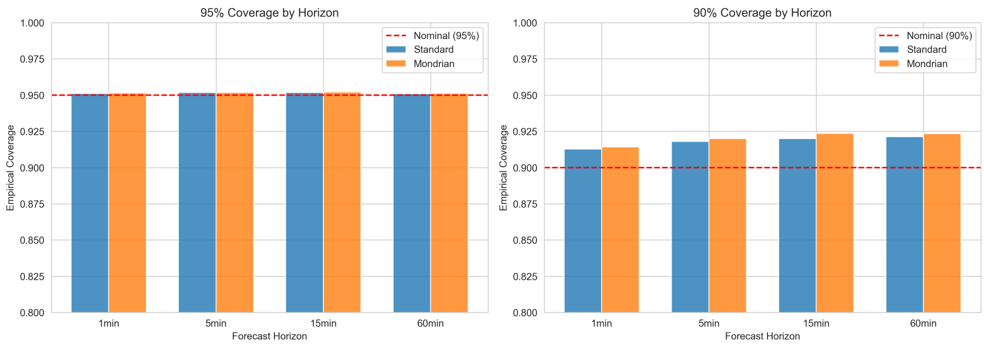

Fig. 2 Empirical coverage by horizon.

Finding 5 (Efficiency / Interval Width): Figure 3 plots the average interval width for both methods. The MiQ-MCP (Mondrian) intervals are consistently and marginally wider than the Standard CQR intervals across all horizons and at both 90% and 95% levels. For example, at the 60-minute horizon (95% level), the mean width for Standard CQR is 0.00305, while for MiQ-MCP it is 0.00311.

An analyst who stopped here would falsely conclude that Standard CQR is the superior method, as it appears to provide the same valid marginal coverage (Figure 2) for a lower cost (Figure 3). The next section demonstrates why this conclusion is incorrect and dangerous.

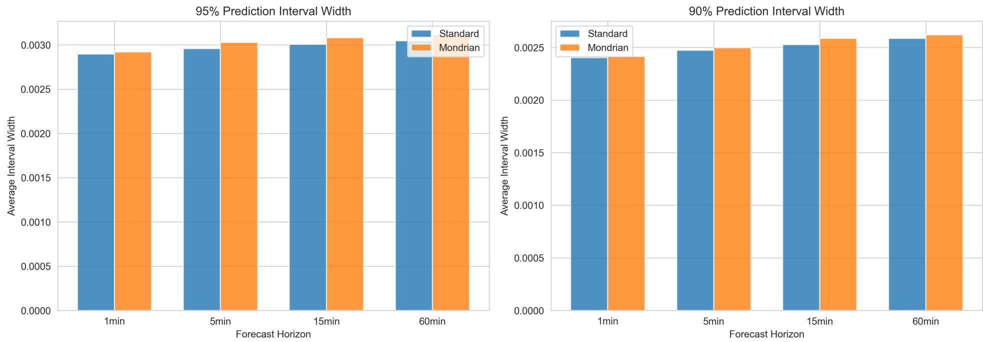

Fig. 3 Average prediction interval width by horizon.

5.3 Conditional Coverage by Time-of-Day (Primary Finding)

This is the central finding of this paper. The "average" performance in Section 5.2 masks a systematic, conditional failure.
Finding 6 (Conditional Failure of Standard CQR): Figure 4 plots the 95% empirical coverage conditionally, broken down by the 10 time-of-day buckets (B0-B9), for the 1-minute horizon. The Standard CQR method (blue bars) exhibits severe conditional failure.

- Under-coverage: In buckets B0-B2 (representing the volatile market open), empirical coverage drops significantly below the 95% nominal line.
- Over-coverage: In buckets B4-B8 (representing the quiet midday), coverage climbs well above 95%.

This is the precise failure we hypothesized. The Standard CQR method uses a single global calibration term $\hat{q}$ (from Algorithm 1), which is an average across all times of day. This $\hat{q}$ is too small for the high-volatility open (leading to narrow, over-confident intervals that under-cover) and too large for the low-volatility midday (leading to wide, inefficient intervals that over-cover).

Finding 7 (Conditional Robustness of MiQ-MCP): Figure 4 clearly shows that the MiQ-MCP (Mondrian) method (orange bars) solves this problem. The orange bars remain remarkably stable, tracking the 95% nominal dashed line almost perfectly across all 10 time buckets. This is a clear visualization of the group-conditional validity provided by the Mondrian (MCP) approach [10]. By calculating a separate $\hat{q}^{(g)}$ for each time bucket g (from Algorithm 2), MiQ-MCP correctly applies a larger calibration adjustment for buckets B0-B2 and a smaller one for B4-B8. It successfully adapts its calibration to the predictable intraday volatility structure.

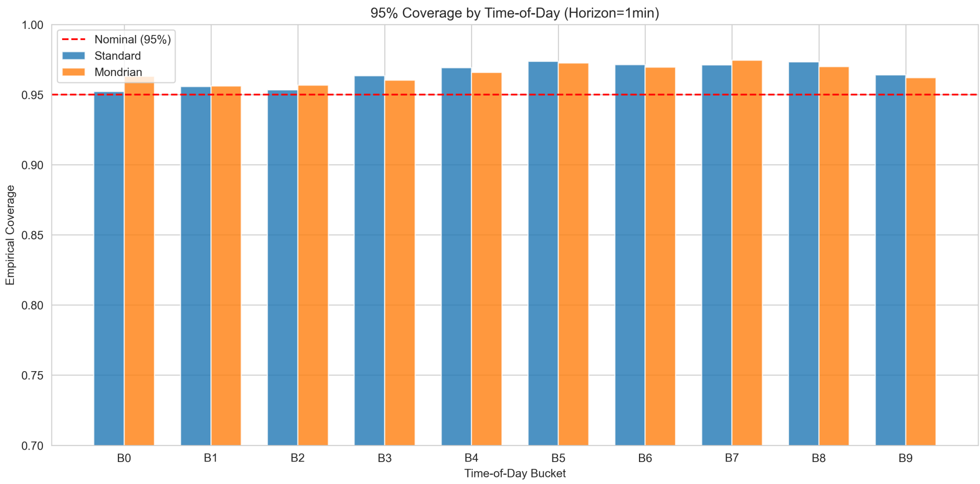

Fig. 4 95% Coverage by time-of-day.

This result is definitive. The "efficiency" of the Standard CQR model (seen in Figure 3) was an illusion, achieved by systematically failing to provide adequate coverage during the most volatile, high-risk periods. MiQ-MCP provides true, robust coverage at all times of day. Its slightly wider average width (Figure 3) is simply the necessary "cost of robustness" required to correctly cover the high-volatility periods that the standard method missed.

6 Discussions

6.1 Interpretation of Findings

Our results demonstrate that for high-frequency financial data, marginal coverage guarantees are not just insufficient; they are actively misleading. They provide a false sense of security by averaging away systematic, conditional failures. The systematic under-coverage of Standard CQR at the market open (shown in Figure 4) is a critical flaw for any practical risk system, as it fails precisely when risk is highest.

Our proposed MiQ-MCP, by integrating a simple, time-of-day stratification [10] into the CQR framework, effectively corrects for this strong intraday seasonality. It delivers the robust, group-conditional coverage that practitioners require. This confirms that the Mondrian CP framework [8] is a highly effective and computationally cheap tool for handling predictable, periodic non-stationarity.

6.2 The Efficiency vs. Conditional Validity Trade-off

The finding that MiQ-MCP intervals are, on average, wider (Figure 3) must be addressed. This is not a failure of the model, but a quantifiable price of robustness. The Standard CQR model's "narrower" intervals were not a sign of efficiency, but a symptom of its failure to widen sufficiently at the market open. MiQ-MCP's "wider" average width is a direct and necessary consequence of correctly widening the intervals in high-risk periods (e.g., B0-B2), a trade-off that is essential for any prudent risk management system.

6.3 Limitations

The primary limitation of our proposed MiQ-MCP is its static nature. The model relies on a single, fixed train-calibration-test split [9]. While we have proven its robustness to intraday seasonality, it is not designed to be robust to inter-day distribution shifts [16]. Financial markets are non-stationary [2]; a model calibrated on data from a low-volatility regime (e.g., 2017) will likely fail (under-cover) during a sudden market crash (e.g., 2020). Its calibration quantiles, $\hat{q}^{(g)}$, are fixed and do not adapt to new market data as it arrives.

7 Conclusion & Future Works

In this paper, we introduced MiQ-MCP, a novel framework for robust uncertainty quantification in high-frequency financial time series. By integrating a lightweight linear quantile regressor with Fourier features to model intraday seasonality and applying Mondrian Conformal Prediction via time-of-day stratification, MiQ-MCP achieves valid marginal coverage while ensuring group-conditional robustness against the U-shaped volatility pattern. Empirical results on 10 years of NIFTY-50 data confirm its superior conditional validity over standard CQR, with statistically significant point forecast improvements (Diebold-Mariano test, $p < 0.05$) and only a minor efficiency trade-off in interval width.

Future work should focus on transitioning from static to adaptive conformal prediction to handle large-scale, inter-day regime shifts in financial time series. A promising direction is integrating Adaptive Conformal Inference or its extensions such as AgACI with our Mondrian framework, yielding an Adaptive Mondrian Conformal Predictor (AMCP) that maintains separate, recursively updated miscoverage targets $\alpha_t^{(g)}$ for each time-of-day bucket [28]. Further improvements include defining Mondrian categories dynamically—e.g., based on recent volatility estimates (EWMA $\hat{\sigma}_t$) rather than fixed clock times—to better capture anomalous regimes such as midday flash crashes. Finally, extending the approach to the multivariate setting would enable the construction of joint prediction regions for portfolios that explicitly account for time-varying asset co-dependencies, moving beyond the current univariate per-stock focus.

Declarations

- Conflict of interest/Competing interests: The authors declare no conflicts of interest.

- Data availability: Data is publicly available from NSE/BSE downloads page.
- Code availability: Github link is available in the references section.
- Author contribution: Author contribution: Conceptualization, SB, RD, DS, AC; methodology, SB, RD; software, SB, RD; validation, SB, RD, DS, AC; formal analysis, SB, RD; investigation, SB, RD; resources, SB, RD, DS, AC; data curation, SB, RD; writing—original draft preparation, SB, RD, DS, AC; writing—review and editing, SB, RD, DS, AC, AW; visualization, SB, RD; supervision, DS, AC; project administration, DS, AC. All authors have read and agreed to the published version of the manuscript.

Appendix A

A.1 Full per-stock DM test (1-min)

The full per-stock DM test for 1 minute time interval is shown at Table A1.

Table A1 Appendix: Full per-stock DM test (1-min Horizon)

Stock	MiQ–MCP MAE	Naive MAE	DM Statistic	DM p-value	Sig. (p<.05)
BEL	0.000682	0.000668	10.027	0.00e+00	Yes
TATASTEEL	0.000568	0.000556	10.458	0.00e+00	Yes
SUNPHARMA	0.000447	0.000448	-7.946	2.00e-15	Yes
GRASIM	0.000536	0.000538	-7.572	3.69e-14	Yes
HINDUNILVR	0.000390	0.000391	-6.062	1.34e-09	Yes
APOLLOHOSP	0.000483	0.000485	-5.136	2.80e-07	Yes
AXISBANK	0.000493	0.000493	-5.107	3.28e-07	Yes
ADANIENT	0.000679	0.000673	5.092	3.54e-07	Yes
ULTRACEMCO	0.000460	0.000461	-4.979	6.40e-07	Yes
TECHM	0.000540	0.000542	-4.974	6.57e-07	Yes
DRREDDY	0.000460	0.000461	-4.863	1.16e-06	Yes
EICHERMOT	0.000526	0.000528	-4.225	2.39e-05	Yes
RELIANCE	0.000390	0.000390	-4.037	5.42e-05	Yes
NESTLEIND	0.000444	0.000444	-3.605	3.12e-04	Yes
BHARTIARTL	0.000473	0.000474	-3.564	3.65e-04	Yes
ASIANPAINT	0.000421	0.000422	-3.435	5.93e-04	Yes
NTPC	0.000593	0.000590	3.258	1.12e-03	Yes
MARUTI	0.000440	0.000441	-3.220	1.28e-03	Yes
CIPLA	0.000496	0.000497	-3.063	2.19e-03	Yes
JSWSTEEL	0.000554	0.000554	-3.041	2.36e-03	Yes
KOTAKBANK	0.000423	0.000424	-2.929	3.41e-03	Yes
POWERGRID	0.000578	0.000572	2.736	6.23e-03	Yes
HEROMOTOCO	0.000546	0.000547	-2.572	1.01e-02	Yes
INDUSINDBK	0.000521	0.000522	-2.520	1.17e-02	Yes
COALINDIA	0.000630	0.000628	2.429	1.51e-02	Yes
HDFCBANK	0.000372	0.000372	-2.241	2.50e-02	Yes
ADANIPORTS	0.000652	0.000653	-2.120	3.40e-02	Yes
HCLTECH	0.000481	0.000482	-1.776	7.58e-02	No
MM	0.000571	0.000571	-1.763	7.78e-02	No
SHRIRAMFIN	0.000685	0.000685	-1.707	8.79e-02	No
JIOFIN	0.000570	0.000566	1.436	1.51e-01	No
BAJAJ-AUTO	0.000493	0.000494	-1.105	2.69e-01	No
LT	0.000462	0.000462	-1.077	2.81e-01	No
WIPRO	0.000521	0.000516	0.983	3.26e-01	No
TATAMOTORS	0.000512	0.000512	0.980	3.27e-01	No
ONGC	0.000600	0.000596	0.908	3.64e-01	No
TITAN	0.000470	0.000471	-0.790	4.30e-01	No
TATACONSUM	0.000515	0.000515	0.418	6.76e-01	No
HINDALCO	0.000589	0.000588	-0.402	6.88e-01	No

References

[1] Leeuwen, P.J., Chiu, J.C., Yang, C.K.: Uncertainty quantification for deep learning. arXiv preprint arXiv:2405.20550 (2024)

[2] Candès, E.J.: Conformal inference when data is not exchangeable. Seminar, Department of Statistics, Stanford University. Joint work with Isaac Gibbs, Rina Foygel Barber, Aaditya Ramdas, and Ryan J. Tibshirani (2022). https://statistics.stanford.edu/events/conformal-inference-when-data-not-exchangeable

[3] Sun, S., Yu, R.: Conformal prediction for time-series forecasting with change points. arXiv preprint arXiv:2509.02844 (2025)

[4] Oliveira, R.I., Orenstein, P., Ramos, T., Romano, J.V.: Split conformal prediction and non-exchangeable data. Journal of Machine Learning Research **25**(225), 1–38 (2024)

[5] Romano, Y., Patterson, E., Candes, E.: Conformalized quantile regression. Advances in neural information processing systems **32** (2019)

[6] developers: FourierFeatures: Fourier Features for Time Series Seasonality. sktime Project, (2025). sktime Project. Documentation. https://www.sktime.net/en/stable/api_reference/auto_generated/sktime.transformations.series.fourier.FourierFeatures.html

[7] Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice (3rd Ed.), pp. 462–466. OTexts, ??? (2021)

[8] Fontana, M., Zeni, G., Vantini, S.: Conformal prediction: a unified review of theory and new challenges. Bernoulli **29**(1), 1–23 (2023)

[9] Baitalik, S.: MiQ-MCP. GitHub repository (2025). https://github.com/sanjanbaitalik/MiQ-MCP.git

[10] Developers, M.: "Theoretical Description – Mondrian Conformal Prediction", v0.9.0 edn. Quantmetry / MAPIE Project, (2022). Quantmetry / MAPIE Project. Accessed: 2025-11-06. https://mapie.readthedocs.io/en/v0.9.0/theoretical_description_mondrian.html

[11] Zhang, C., Zhang, Y., Cucuringu, M., Qian, Z.: Volatility forecasting with machine learning and intraday commonality. Journal of Financial Econometrics **22**(2), 492–530 (2024)

[12] Moreno-Pino, F., Zohren, S.: Deepvol: Volatility forecasting from high-frequency data with dilated causal convolutions. Quantitative Finance **24**(8), 1105–1127 (2024)

[13] Lee, J., Xu, C., Xie, Y.: Kernel-based optimally weighted conformal time-series prediction. In: The Thirteenth International Conference on Learning Representations (2025). https://openreview.net/forum?id=oP7arLOWix

[14] Lee, J., Xu, C., Xie, Y.: Conformal prediction for time series with transformer. In: ICML 2024 Workshop on Structured Probabilistic Inference & Generative Modeling (2024). https://openreview.net/forum?id=3dDDKaSrye

[15] Zhang, X., Kang, Y., Li, C., Wang, W., Wang, K.: Lstm-conformal forecasting-based bitcoin forecasting method for enhancing reliability. PLoS One **20**(5), 0319008 (2025)

[16] Barber, R.F., Candes, E.J., Ramdas, A., Tibshirani, R.J.: Conformal prediction beyond exchangeability. The Annals of Statistics **51**(2), 816–845 (2023)

[17] Auer, A., Gauch, M., Klotz, D., Hochreiter, S.: Conformal prediction for time series with modern hopfield networks. Advances in neural information processing systems **36**, 56027–56074 (2023)

[18] Xu, C., Xie, Y.: Sequential predictive conformal inference for time series. In: International Conference on Machine Learning, pp. 38707–38727 (2023). PMLR

[19] Zhao, Z., Liu, H., Prakash, B.A.: Navigating Concept Drift and Temporal Shift: Distribution Shift Generalized Time-Series Forecasting (2025). https://openreview.net/forum?id=Klx0Rq9vbC

[20] Angelopoulos, A., Candes, E., Tibshirani, R.J.: Conformal pid control for time series prediction. Advances in neural information processing systems **36**, 23047–23074 (2023)

[21] Li, R., Rodríguez, A.: Neural conformal control for time series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, pp. 18439–18447 (2025)

[22] Prinzhorn, D.W., Nijdam, T., Linden, P.A., Timans, A.: Conformal time series decomposition with component-wise exchangeability. arXiv preprint arXiv:2406.16766 (2024)

[23] Kaur, J.N., Jordan, M.I., Alaa, A.: Conformal prediction sets with improved conditional coverage using trust scores. arXiv preprint arXiv:2501.10139 (2025)

[24] Sun, S.H., Yu, R.: Copula conformal prediction for multi-step time series prediction. In: The Twelfth International Conference on Learning Representations (2024). https://openreview.net/forum?id=ojIJZDNIBj

[25] Cleaveland, M., Lee, I., Pappas, G.J., Lindemann, L.: Conformal prediction

regions for time series using linear complementarity programming. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 20984–20992 (2024)

[26] Lee, J., Xu, C., Xie, Y.: Flow-based conformal prediction for multi-dimensional time series. arXiv preprint arXiv:2502.05709 (2025)

[27] Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. Journal of Business & economic statistics **20**(1), 134–144 (2002)

[28] Wang, H., Ji, Q.: Epistemic uncertainty quantification for pre-trained neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11052–11061 (2024)