Asymmetric and Self-Adaptive Conformal Inference for Time-Series Forecasting with Distribution Shift

Albert Hsin-Kai Lin and Hsuan-Tien Lin

Computer Science and Information Engineering, National Taiwan University, Taiwan

Abstract. Uncertainty quantification in time-series forecasting holds significant practical importance across various domains. Conformal inference, by providing narrow intervals to cover the true outcomes, stands out as a family of widely used tools in this context. Nevertheless, existing conformal inference tools suffer from two issues. Firstly, they often assume symmetrical intervals around the predicted value, resulting in wide and unrealistic intervals for downstream applications. Secondly, these tools exhibit sensitivity to adaptation parameters in distribution-shifting scenarios. In response to these issues, we introduce a novel method called Asymmetric and Self-Adaptive Conformal Inference (ASACI). ASACI models upper and lower half-intervals separately. Moreover, ASACI uses an internal bandit solver for each half-interval to adapt its parameters autonomously. Beyond its theoretical foundation, we validate the efficacy of ASACI across various common time-series forecasting datasets. We observe that ASACI can generate asymmetric intervals with both valid coverage and realistic ranges across various datasets. Those inherent strengths of ASACI position it as a robust practical solution for incorporating uncertainty quantification in time-series forecasting.

Keywords: conformal inference · time-series · asymmetric interval

1 Introduction

The swift development and adoption of machine learning algorithms in time-series forecasting established the significance of uncertainty quantification [14, 17, 2]. Understanding the certainty of model outputs aids experts in the decision-making process. Notably, uncertainty quantification in time-series forecasting, such as weather or price forecasting, often encounters distributional shifts. Using the VIX Index (representing the 30-day expected volatility of the U.S. stock market) shown in Fig 1 and the NASA Global Land-Ocean Temperature Index depicted in Fig 2 as illustrations, we can discern distinct patterns. The VIX Index displays instances of escalating local volatility at certain intervals, marked by blue shading. Meanwhile, the NASA Global Temperature Index demonstrates shifts in the trend direction, highlighted by red arrows. Each time-series dataset may manifest its unique distributional shifts. Hence, devising an uncertainty

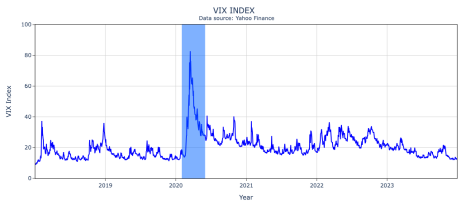

Fig. 1. Local Volatility: VIX Index

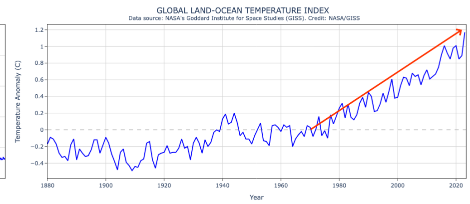

Fig. 2. Trend Changes: Temperature Index

quantification method that adapts to the nature of shifting within the specific time-series data is important.

Conformal inference [35, 24, 20, 37, 13] is a family of methods that solves the uncertainty quantification task with minimal adjustments to the original prediction pipeline. The family generates empirically effective uncertainty intervals and offers a distribution-free guarantee that the intervals cover the true values. Initially designed for more stable distributions, a recent trajectory of research focuses on online settings where data arrive sequentially, such as time-series data. With the introduction of Adaptive Conformal Inference and subsequent advancements [15, 16, 44, 9], conformal inference is able to address distributional shifts in time-series data.

Current adaptive conformal inference techniques primarily focus on generating symmetric uncertainty intervals around predicted values, assuming the error distribution to be symmetric. However, this symmetric approach can lead to overly optimistic estimates on one end and overly conservative estimates on the other. In certain real-world applications, asymmetry is employed to prioritize preferences for over or underestimation, suggesting that uncertainty intervals should not always be symmetric. Another unaddressed challenge with current conformal inference techniques is their sensitivity to the step-size parameter for updating the interval. While some aggregated approaches attempt to mitigate this issue [44, 9], they come at the expense of increased computational complexity. Practical applications demand an efficient solution to manage this sensitivity effectively.

This study introduces Asymmetric and Self-Adaptive Conformal Inference (ASACI) as a solution to the challenges outlined. Our primary contribution lies in offering a practical solution for conformal inference in time-series forecasting. By separately monitoring upper and lower intervals, we achieve tight and valid intervals irrespective of the underlying distribution, regardless of symmetry or asymmetry. Our approach is supported by theoretical guarantees and extensive experiments. Moreover, we introduce bandit solvers as a self-adaptive mechanism to mitigate parameter sensitivity, thereby enhancing algorithm robustness and flexibility. We anticipate that our method will not only advance conformal inference development but also serve as a robust option for integrating conformal prediction into decision-making and other downstream tasks in time-series forecasting.

2 Problem Formulation

2.1 Problem Statement

Consider online time-series forecasting. Suppose we have covariates $x_t \in \mathbb{X}$ and label $y_t \in \mathbb{Y}$, we shall make a forecast $\hat{y}_t$ of y_t. At time steps $t \in \mathbb{N}$, we have the information of $\{(x_i, y_i)_{i\in[0,t)} \cup x_t\}$, and we, besides the point estimation $\hat{y}_t$, want to find a conformal set C_t to cover y_t. The conformal set C_t is desired to be able to cover y_t most of the time, meanwhile C_t should remain as compact as possible. We can choose a target conformal level $\alpha \in (0,1)$ beforehand, which specifies the allowable error rate ($y_t \notin C_t$). Ideally, we attempt to achieve target coverage, namely $y_t \in C_t$ for at least $(1-\alpha)\%$ times on average.

$$\lim_{N\to\infty} \left(\frac{\sum_N \mathbb{1}_{y_t \notin C_t}}{N}\right) = 1 - \alpha$$

To construct such C_t, we can find a suitable $\hat{s}$ with $C_t = [\hat{y} - \hat{s}, \hat{y} + \hat{s}]$. To find such $\hat{s}$, we can choose a score function S, where $S(\hat{y}_t, y_t)$ measures the accuracy of our prediction $\hat{y}_t$. Then, we model the conformal set C_t by finding a suitable quantile of such $\{S_{1:t-1}\}$. By selecting appropriate score functions S, we can build a conformal set that fits the different nature of the data distribution.

2.2 Relaxing Exchangeability

Exchangeability is a concept that generalizes the notion of identically distributed random variables without assuming independence. In the convention line of conformal inference research, exchangeability plays a large part in the assumption. However, it is often not the case in time-series data. There are different attempted strategies to handle non-exchangeable time series. Two lines of strategies are weighted conformal prediction [32] and adaptive conformal prediction [15]. Adaptive conformal inference (referred to as ACI below) [15] is an interesting strategy that treats time-series conformal inference as an online learning problem. ACI states since the data distribution will shift over time, the target conformal level α, might be outdated and the C_t formed will not be able to achieve target coverage. Instead, there might be an α^*, used for forming C_t^*, which will be able to achieve an error rate of α. Thus, we shall adaptively update α_t of time step t to approach α^*.

$$\alpha_{t+1} = \alpha_t + \gamma(\alpha - \sum_{s=1}^{t} w_s err_t), \quad err_t = \begin{cases} 1 \text{ , } y_t \notin C_t(\alpha_t) \\ 0 \text{ , } otherwise \end{cases}$$

We can update α_{t+1} after the realization of y_t using the information of err_t, where err_t is the indicator of error or correctness. We shall choose a step size $\gamma > 0$ for updating α_{t+1}. The weight w_s can be a simple one-hot vector at timestep t or even an exponentially weighted function. By doing so, we can approximate $\alpha_t \to \alpha^*$ in the long run.

2.3 Introducing Asymmetry

ACI establishes a firm ground for the adaptiveness of conformal inference. However, in investigating more real-world data, such as stock volatility Fig 3, we found the need to introduce asymmetry into forming a conformal set C_t.

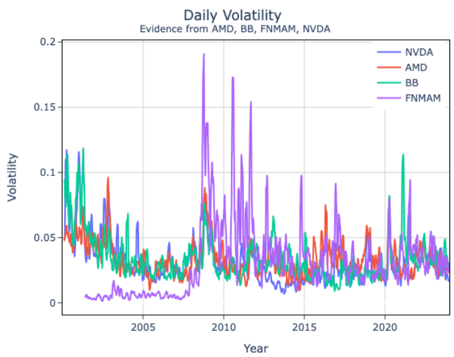

Fig. 3. Daily Stock Volatility

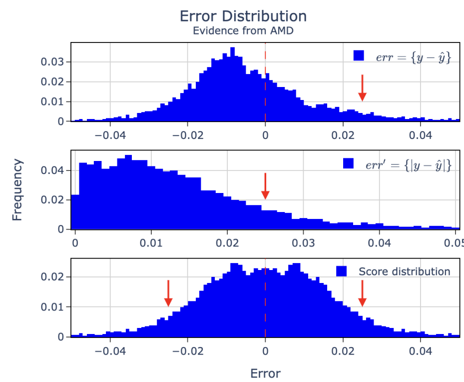

Fig. 4. AMD Error Distribution

ACI uses the full range of $y_{1:t}$ and $\hat{y}_{1:t}$ to calculate score $s_{1:t}$. The conformal set C_t formed will treat the error distribution in a symmetric perspective. However, an example of volatility forecasting on the stock AMD using the GARCH model [10] can illustrate the need to include asymmetry. The error distribution $\{y - \hat{y}\}$ is positively skewed and negatively centered as shown in Fig 4. The figure at the top is the original error distribution. The figure in the middle is the distribution of the score function. The figure at the bottom illustrates the implicit assumption on the error distribution if we take the quantile of the full score distribution, which destroys the original error distribution and turns it into a symmetric distribution. The quantile, represented by the red error for negative error, is unreasonable, which poses an issue in the value range of the prediction set C_t.

In order to address the problem, ASACI split $y_{1:t}$ and $\hat{y}_{1:t}$ into two sets. Namely, the upper error set $E_t^u = \{(y_t, \hat{y}_t); y_t \geq \hat{y}_t\}$ and the lower error set $E_t^l = \{(y_t, \hat{y}_t); y_t \leq \hat{y}_t\}$. We can find the α_u and α_l for the upper and lower bounds of $C_t = [\hat{y} - q(\alpha_l), \hat{y} + q(\alpha_u)]$ for some function q respectively. Thus, we separately model the upper and lower bounds to produce an asymmetric conformal set C_t. It is straightforward that splitting distributions can help produce better intervals when the distribution is asymmetric. In contrast, if the distribution is symmetric, α_u and α_l converge to the same α^* in the long run. Thus, we can achieve an adaptive conformal inference for different kinds of data under the same scheme. More details will be rolled out in the fourth section.

3 Related Work

A recent introduction [4] [30] covers the most popular methods in conformal inference and the development of conformal prediction. The development of Conformal Prediction dates back to the late 90s and the early 2000s by [35]. Followed by several works [24] [20] [37] [13], conformal prediction has attracted a lot of attention. Many earlier works focus on dealing with conformal prediction under stable prediction and given the strong assumption of exchangeability. Starting with Split Conformal [24] [24], it splits the dataset into the training set and calibration set. The former is used to fit the prediction model, while the latter is used to construct the conformal set. An alternative to splitting the dataset is full conformal [24]. Full conformal uses the whole dataset for training and heavily relies on the assumption of exchangeability. For a new data point X_i to be predicted, it will find a hypothesized $y \in \mathbb{Y}$ that follows exchangeability and fits all data along with the hypothesized one. By doing so, it can achieve tight and theoretical guarantees. An obvious drawback is the intense computation; thus, Approximate Full Conformal [21] [22] [1] is a line of work trying to relieve the burden. Also, between the two extremes, Cross-Conformal Prediction [34] [36], CV+ [29], Jackknife+ [7], RASP [3] serves as an attempt to find the balance between full conformal and split conformal. However, these works do not perform well beyond exchangeability. Thus, they may not handle time-series data effectively.

Following the line of the works, many attempts deal with distributional shifts in conformal prediction. For example, EnbPI [39], EnCQR [18], and ERAPS [40] link ensemble methods and conformal prediction. However, the computational cost of the ensemble method might not be usable in some scenarios where efficiency is crucial. Stepping into dealing with the problem of distributional shift, we have weighted split conformal [32] and conformal beyond exchangeability [8], which is an attempt to relieve exchangeability through re-weighting. The core idea is to believe that the nearer error should have a more significant meaning than those that occurred long ago. More recent works [23] [25] [26] [27] [43] [31] [6] [42] [41] also relax the assumption of data exchangeability and propose more methods for forming robust conformal sets on time series problems. However, these methods might not be as lightweight as adaptive conformal inference is, and all their methods have their assumptions.

Closer lines of our work start with Adaptive Conformal Inference [15] and are followed by several works [16] [44] [9] [38]. The latter work expands the former either by introducing aggregated methods to enhance stability or by enhancing the theoretical guarantee. The aggregation method, similar to the ensemble method, requires maintaining several ACI, which might increase computational cost. A recent work, Conformal PID [5], also provides a more general framework to deal with adaptive conformal inference. The work directly predicts the quantile, aiming to resolve extreme cases where the error is large. However, predicting the quantile itself requires careful tuning and additional overhead. This line of work is promising and effective, but the lack of dealing with asymmetry is a gap toward real-life adoption. Another work of Conformal Quantile Regression

[28] modeled the quantile asymmetrically by using quantile regression to directly output quantiles for the conformal set. However, they lack dealing with online distributional shifts and focus on exchangeable settings. Thus, our work, ASACI, combining the merits of adaptiveness and asymmetry, can set down a practical and realistic method for adopting conformal prediction, regardless of the data distribution.

4 Method: ASACI

In order to generate a realistic and valid interval range, we shall take into account the fact that the error distribution might be asymmetric or skewed due to the nature of the data. Thus, ASACI introduces asymmetry to form the conformal set C_t. After introducing asymmetry, we still need to deal with the problem of sensitivity in step size γ. Thus, by incorporating bandit, we can solve the mentioned problems simultaneously.

The details of the algorithm are depicted below. First, we will form two error distributions $E_t^u = \{(y_t, \hat{y}_t); y_t \geq \hat{y}_t\}$ and $E_t^l = \{(y_t, \hat{y}_t); y_t \leq \hat{y}_t\}$. Later, we will calculate the score distribution respectively, namely $S_{1:t}^i = S(y_t, \hat{y}_t); (y_t, \hat{y}_t) \in E_t^i$ for $i \in \{u, v\}$. Later, given 2 score distributions, we retrieve the quantile $q_t^i = Q_{1-\alpha_t^i}(S_t^i)$ for $i \in \{u, v\}$. Thus, we can form the asymmetric conformal set $C_t = [\hat{y}_{t+1} - q_t^l, \hat{y}_{t+1} + q_t^u]$.

Later, we can consider three scenarios for updating α_{t+1}^i. First, if $y_{t+1} > \hat{y}_{t+1}$ and $y_t \notin C_t$, it is straightforward that we shall expand the upper bound, which is done by making $\alpha_{t+1}^u < \alpha_t^u$. However, for α_t^l, whether increasing or decreasing, is a choice between exploration and exploitation. Namely, $y_{t+1} > \hat{y}_{t+1}$ and $y_t \notin C_t$ might represent a general increase in the scale of variance, or it can simply mean a temporary shift of the error quantile toward the positive axis. To deal with the issue, we introduce a bandit in deciding the direction and step size γ_{t+1}^l in updating α_{t+1}^l. Thus, we will form a set of actions of γ_{t+1}^l, such as $\{10, 1, 0.1, 0, -1\}$, which contains both positive and negative step sizes with different magnitudes that represent decreasing and increasing the α_t^l respectively.

Next, for $y_{t+1} \in C_t$, we will tighten the bound and increase the α_{t+1}^i for $i \in \{u, v\}$. Bandit can thus help to decide the step size in updating the α_{t+1}^i, which increases the parameter-agnostic ability of the whole method.

4.1 Algorithm Walkthrough

The algorithm for the Asymmetric and Self-Adaptive Conformal Inference is provided below Algo 1. It is composed of 2 bandits B_i with i indicating we are updating the upper or lower α_t^i. We shall sample an action b_i from B_i, which is the step size chosen by the bandit. Then, we can get the weight $w_i = \gamma b_i$ for updating our α_t^i. The intuition behind the design is that each bandit will be separately balanced between exploration and exploitation. By doing so, we can better differentiate a temporary shift of our error distribution from an increase in error variance.

Algorithm 1 ASACI

INPUT: $\alpha \in (0,1)$, $\gamma > 0$, $i \in \{U, L\}$

while t in 1,2 ... N **do**
  $E^i_{1:t} \leftarrow \{(y_j, \hat{y}_j; y_j \geq \ or \ \leq \hat{y}_j)\}\ with\ j \in 1 \ldots t$
  $S^i_{1:t} \leftarrow S(y_t, \hat{y}_t)\ for\ (y_t, \hat{y}_t) \in E^i_{1:t}$
  $q^i_t \leftarrow Q_{1-\alpha^i_t}(S^i_{1:t})$
  $predict\ \hat{y}_{t+1}\ with\ \{x_{1:t+1}, y_{1:t}\}$
  $C_t := [\hat{y}_{t+1} - q^l_t, \hat{y}_{t+1} + q^u_t]$
  $\alpha^u_{t+1} \leftarrow \alpha^u_t + b_U * \gamma(\alpha - err_t)$
  $\alpha^l_{t+1} \leftarrow \alpha^u_t + b_L * \gamma(\alpha - err_t)$
  Update Reward for B_U, B_L
end while =0

The reward design for the bandit impacts the algorithm's effectiveness. We aim for a reward that effectively captures trends. Specifically, as the error grows larger, we should adopt a more aggressive approach in updating α^i_t. However, we should avoid excessive updates if the trend of error growth appears to decelerate. The following formula gives a simple but effective design used in the paper where $\bar{e}_t = \frac{1}{10}\sum_{i=t-9}^{i} e_i$ and $\Delta e = e_t - \bar{e}_t$. Also, $\kappa = 1$ and $\beta = 0.1$ are hyperparameters that control the sensitivity of the reward term.

$$r_t = \begin{cases} \kappa \cdot a_t \cdot (1 + \bar{e}_t) + \beta \cdot \left(0 \text{ if } a_t = 0 \text{ else } \frac{\text{sign}(\Delta e_t)}{|a_t|}\right), & \text{if } \bar{e}_t > \alpha, \\ -\kappa \cdot a_t \cdot (1 + \bar{e}_t) - \beta \cdot \left(0 \text{ if } a_t = 0 \text{ else } \frac{\text{sign}(\Delta e_t)}{|a_t|}\right), & \text{if } \bar{e}_t \leq \alpha. \end{cases}$$

However, to mitigate the effect of different scaling, we will apply a min-max scale to the reward function and normalize it between $[0, 1]$. It is helpful in stabilizing the value range of our bandit reward. The normalized $\tilde{r_t}$ is then used to update the bandit.

$$\tilde{r_t} = \frac{r_t - \min_{i<t} r_i}{\max_{j<t} r_j - \min_{i<t} r_i}$$

An algorithm that will be used to compare our experiment results is the asymmetric ACI. It is a simplified version of ASACI that replaces the bandit with a fixed-size step ($b_U = b_L = b$) to serve as a comparison.

5 Theoretical Guarantee

The theoretical guarantee includes two parts. First, we will look at how the algorithm holds under exchangeability. Second, we will walk through how the introduction of asymmetry and bandit solver maintains theoretical guarantees.

Assume that our target conformal level is α and step size γ. At each time step t, we have our adaptive alpha α^i_t and sampled bandit action b^i_t for $i \in \{U, L\}$, representing upper and lower error modeling, respectively. We then generate quantile q_t^i from error set E_t^i with $Q(1-\alpha_t^i; E_t^i)$. Especially, we have a model $\hat{f}$ trained on some set of data, and then we perform ASACI on the testing set that is disjoint from the training set.

Proposition 1 *Under exchangeability, suppose that* $lim_{t\to\infty}\frac{t}{t'} = k$ *with* $t' = \|\{Y_i; Y_i - \hat{f}(X_i) > 0\}\|$, *we have* $q_t^u = Quantile(\lceil(1-\alpha)(t+1)k\rceil; E_t^u)$ *such that* $P(Y_{t+1} \in C_t) \geq 1-\alpha$.

Proof. For any score function $S(\cdot)$ which is order-preserving with $\{y_i - \hat{f}(x_i)\}$, we have $E_t = \{s_1, s_2, \dots s_t\}$. We know that with exchangeability, we have

$$q_t^u = Quantile(\lceil(1-\alpha)(t+1)\rceil; E_t)$$

such that $P(Y_{t+1} \in (-\infty, q_t^u]) \geq 1-\alpha$ as held by [24].
For a subset $E_t^u = \{s_i; s_i > 0\}$ of E_t, we have $\|E_t^u\| = t'$. Thus,

$$\begin{aligned} q_t^u &= Quantile(\lceil(1-\alpha)(t+1)\rceil; E_t) \\ &= Quantile(\lceil(1-\alpha)(t'+1)\frac{t+1}{t'+1}\rceil; E_t^u) \end{aligned}$$

For $t \to \infty$, we know that $\frac{t+1}{t'+1} \to k$, thus we have

$$q_t^u = Quantile(\lceil(1-\alpha)(t'+1)k\rceil; E_t^u)$$

Thus, $P(Y_{t+1} \in (-\infty, q_t^u]) \geq 1-\alpha$ holds.

Proposition 2 *Suppose that the true* $\alpha^* \in [-k, k]$ *and the range of bandit action* $b_t^i \in [-s, s]$ *with optimal action* b^*, *then we have*

$$\frac{1}{T'}|\sum_{t=k}^{T}(\alpha - err_t)| \leq \frac{\gamma s + \max\{1-\alpha, \alpha\}}{T'\gamma b^*}$$

for $\exists k$ *such that* $b_t^i \to b^*$ *and define* $T' = T - k$.

Proof. Given that the true $\alpha^{*,i}$ is bounded in some interval $[-k, k]$, thus for α_{t+1}^i, we have

$$\alpha_{t+1}^i = \alpha_0^i + \sum_{j=0}^{T}\gamma b_j^i(\alpha - err_j) \Rightarrow \frac{\alpha_{t+1}^i - \alpha_0^i}{T\gamma s} = \frac{1}{T}\sum_{j=0}^{T}\frac{b_j^i}{s}(\alpha - err_j)$$

As $T \to \infty$, we have k with $b^i_{j \geq k} \to b^*$, where the bandit found the optimal choice,

$$\because \alpha^i_{t+1} \in [-\gamma s, 1+\gamma s] \Rightarrow |\alpha^i_{t+1} - \alpha^i_0| \leq \max\{\alpha, 1-\alpha\} + \gamma s$$

$$\therefore \frac{\max\{\alpha, 1-\alpha\} + \gamma s}{T\gamma s} \geq \frac{1}{T}|\sum_{j=0}^{T} \frac{b^i_j}{s}(\alpha - err_j)|$$

$$\Rightarrow \frac{\max\{\alpha, 1-\alpha\} + \gamma s}{T'\gamma s} \geq \frac{1}{T'}|\sum_{j=k}^{T} \frac{b^*}{s}(\alpha - err_j)|$$

$$\Rightarrow \frac{\max\{\alpha, 1-\alpha\} + \gamma s}{T'\gamma b^*} \geq \frac{1}{T'}|\sum_{j=k}^{T} (\alpha - err_j)|$$

Also, the coverage relies on the convergence of the bandit toward the optimal action in reasonable steps.

6 Experiment

We tested our algorithm against several datasets. Our main metric for the experiment is to achieve the target coverage with a possibly small width of the conformal set. Also, since we are introducing asymmetry into the formation of the conformal set, we wish to beat other methods in the dataset that exhibit asymmetry or skewness, and those with asymmetric loss functions. On other datasets, we aim to achieve a similar or even better level of coverage and width.

6.1 Datasets

The first dataset we use is market volatility for AMD, BB, FNMAM, and NVDA. Since market volatility is skewed, we highlight the importance of asymmetric intervals. The second dataset is the federal fund rate dataset [1] and the Beijing PM2.5 dataset [11], where we apply an asymmetric loss function to present the effectiveness of our method. Last, we focus on more general datasets used in time-series regression tasks – ETT dataset [46], traffic [2], exchange rate [19], electricity [33], and the weather dataset [3]. The goal of these datasets is to keep up with or even beat SOTA.

6.2 Metrics

First, we should measure the ability of a set C_t to cover the true label y_t under α percentage. Thus, we define the local coverage $LocalCov_t$ at time step t,

$$LocalCov_t = 1 - \frac{1}{n} \sum_{i=t-n}^{t} 1_{y_i \in C_{i-1}}$$

[1] Federal Funds Effective Rate: https://fred.stlouisfed.org/series/DFF
[2] Traffic Dataset: https://pems.dot.ca.gov/
[3] Weather Dataset: https://www.bgc-jena.mpg.de/wetter/

When we are taking a target conformal level α, we will aim for a local coverage of $1-\alpha$. There might be some other auxiliary metric, such as the volatility of the local coverage, for which smaller values are better. However, we will only focus on the average of local coverage for now.

Another important metric is the width of the conformal set. Although a smaller width doesn't necessarily mean a better or more valid conformal set, it is generally a good indicator of the quality of the conformal set.

$$AvgWidth = \sum_{t=0}^{n} ||C_t||$$

The last indicator, which is specific to the asymmetric case we mentioned, would be the number of occurrences of invalid conformal sets, that is, sets containing a negative range. In other datasets, it is harder to evaluate the data range, but for the case of stock volatility, federal fund rate forecasting, and pm2.5 should shed light on the idea of the validity of intervals.

$$Invalid = \frac{1}{n} \sum_{i=t-n}^{t} 1_{(-\infty,0) \cap C_{i-1}}$$

6.3 Asymmetry Distribution – Stock Volatility Prediction

When it comes to applying time-series forecasting to real-world scenarios, it is extremely important to ensure the output of the machine learning algorithm is valid and reasonable. For example, in predicting the weather for an airline company, frequently outputting an invalid value range on weather might cause an unnecessary burden on airplane flight planning. However, it might not be easy to define what kinds of output of weather forecasts are considered invalid. Thankfully, in volatility forecasting, it is easy to know whether a forecast's range is reasonable or not; that is, volatility should be positive. Thus, we will take stock volatility as an example to demonstrate the ability of ASACI to produce a valid conformal set.

The data of stock price is collected on Yahoo Finance [4] and dates from 1990/1/1 (or the earliest available date) to 2023/12/31. We take the future 7 days realized standard deviation of returns as the realized volatility, and apply the GARCH(1,1) model to predict future volatility with rolling past 250 days as training data. The experiment result is presented below. The method with an asterisk (*) is ASACI with different bandit methods, while the R means that the bandit algorithm will only consider the newest $k = 250$ rewards and scores. If the invalid is smaller than the baseline ACI, it will be labeled in bold text, while coverage will be labeled in bold if it successfully achieves the target 90%.

From the experiment Tab1 results of NVDA, AMD, BB, and FNMAM, you can observe that ASACI is able to produce a more valid (greater than 0) conformal set C_t while achieving the targeted conformal level. However, CQR failed to

[4] Yahoo Finance: https://finance.yahoo.com/

	NVDA			AMD		
Method	**Invalid**	**Coverage**	**Width**	**Invalid**	**Coverage**	**Width**
CQR	0.00%	70.63%	0.0297	0.00%	71.15%	0.0361
ACI	12.01%	**89.68%**	0.0485	15.48%	**89.66%**	0.0581
Asym-ACI	**8.21%**	**89.50%**	0.0563	**12.35%**	**89.41%**	0.0653
***UCB1**	**9.22%**	**91.44%**	0.0604	14.67%	**91.42%**	0.0699
***UCB1-R**	12.73%	**89.92%**	0.0575	**13.56%**	**90.32%**	0.0687
***TS**	**5.71%**	**89.17%**	0.0586	**11.60%**	**89.25%**	0.0641
***TS-R**	**7.23%**	**89.00%**	0.0562	**8.63%**	**89.38%**	0.0682
	BB			**FNMAM**		
Method	**Invalid**	**Coverage**	**Width**	**Invalid**	**Coverage**	**Width**
CQR	0.00%	72.14%	0.0365	0.00%	71.45%	0.0506
ACI	17.77%	**89.84%**	0.0629	49.92%	**89.88%**	0.0903
Asym-ACI	**11.33%**	**89.49%**	0.0731	**46.06%**	**89.61%**	0.1064
***UCB1**	**13.40%**	**91.42%**	0.0807	51.17%	**91.08%**	0.1123
***UCB1-R**	**14.88%**	**91.77%**	0.0812	56.75%	**91.75%**	0.117
***TS**	**10.86%**	**89.67%**	0.0748	**46.60%**	**88.87%**	0.1025
***TS-R**	**6.95%**	**88.61%**	0.0781	56.43%	**90.53%**	0.1102

Table 1. Stock Volatility Experiment

produce valid coverage, although it can produce asymmetric intervals. Generally, Thompson Sampling and Thompson Sampling with a rolling reward can achieve the best results in volatility forecasting. However, we should take a closer look at the reason that ASACI generates a larger width. The reason is that since we split the error distributions into two parts, the large impact will naturally stay in effect for longer as the number of elements in each error distribution is smaller. Thus, as you can observe in Fig 5, the green area, which labels the area of the conformal set, tends to stay at a larger range longer for ASACI (the one in the top) than the symmetric ACI (the one in the bottom) does. This behavior is expected as the error distribution is skewed, as evidenced in Fig 6. Thus, ASACI can indeed produce a more realistic conformal set.

6.4 Asymmetry Loss Function – PM2.5 and Federal Fund Rate

However, as we stated above, sometimes in real-world forecasting, we will be inclined to either over- or under-estimate, as it could impose significant risk if we are over-optimistic. This is important for some areas such as risk monitoring, hazard forecasting, or so, as these tasks will preferably try to make their predictions tilted to one side [12].

For Federal Fund Rate, sometimes refered to as EFFR, forecasting, and PM2.5 forecasting, we use XGBoost for forecasting. The data used spans from 1954/7/1 to 2023/12/31. Since the valid value range for the Federal Fund Rate and PM2.5 datasets isn't as straightforward as the stock volatility forecasting, we would use only the average coverage and average conformal set width as our metric here. We will prioritize attaining our target coverage before the width of the sets. Our loss function is given as $l(y, \hat{y}) = (0.4 + 0.2 \cdot 1_{y>\hat{y}})|y - \hat{y}|$ in an asymmetric sense.

The result is presented in the Tab 2, where you can observe that in the case of Federal Fund Rate, only UCB1 and UCB1-R can achieve the targeted interval.

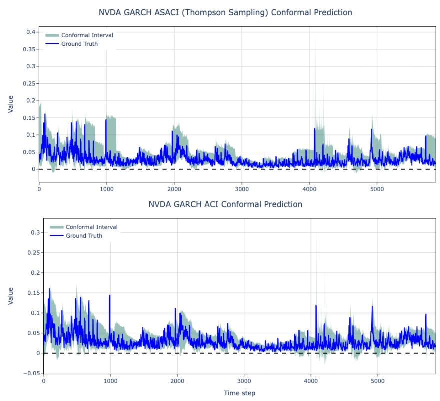

Fig. 5. NVDA Coverage: ASACI vs ACI

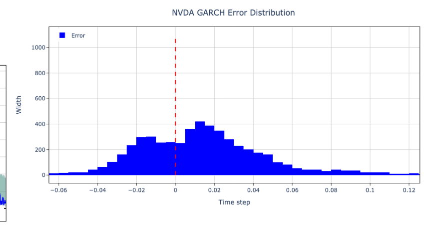

Fig. 6. NVDA Error Distribution

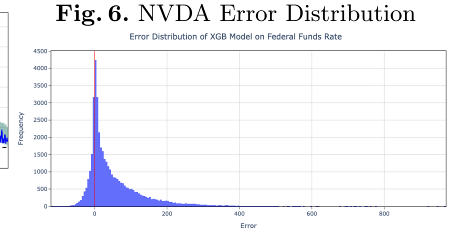

Fig. 7. EFFR Error Distribution

In the case of PM2.5, though ACI and ASACI can achieve targeted intervals, the width of ASACI is smaller than that of ACI. It is also evidenced that the error distribution becomes extremely skewed in Fig 7. Thus, we can conclude that when we adopt an asymmetric loss function, ASACI is a top choice to perform conformal prediction as the nature of the algorithm makes it suit for dealing with adverse against over- or under-estimations.

	Federal Fund		PM2.5	
Method	**Coverage**	**Width**	**Coverage**	**Width**
CQR	64.45%	0.5948	68.40%	97.5
ACI	86.14%	0.7338	**89.51%**	319.2
Asym-ACI	85.44%	0.6840	**89.18%**	**205.2**
***UCB1**	**88.53%**	0.8652	**91.17%**	**213.5**
***UCB1-R**	**88.66%**	0.8494	**90.19%**	**210.3**
***TS**	85.87%	0.7050	84.75%	**171.4**
***TS-R**	84.37%	0.6704	**89.56%**	**209.1**

Table 2. Asymmetric Loss Function Experiment

6.5 Other Dataset

At last, we will examine our method against some of the common datasets to showcase the general ability of ASACI to converge to the targeted conformal level. We will look at the aforementioned 5 datasets predicted by the Transformer model and predict 720 timesteps. The code for inference is from [45].

As you can observe, for the Weather and the Traffic dataset, ASACI can achieve target coverage with a smaller width than the ACI baseline. Most algorithms fall short for the Exchange and Electricity datasets, while ASACI with UCB1 bandit is closest to the target. As for ETTm1, ETTm2, and ETTh2, ASACI with UCB1 bandit can achieve target coverage with a smaller width than the ACI baseline. For ETTh1, the ASACI with the UCB1 bandit can achieve a targeted level with a smaller width than the ACI baseline. Thus, ASACI can often accomplish the targeted conformal level with competitive width.

	Weather		Traffic		Exchange		Electricity	
Method	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width
CQR	96.92%	14.7	68.10%	0.009	45.31%	0.106	94.48%	714.9
ACI	**89.77%**	196.4	**89.32%**	0.042	85.73%	0.336	86.21%	1195.1
Asym-ACI	**89.53%**	**149.4**	**89.71%**	**0.035**	**89.00%**	**0.169**	86.22%	1182.0
*UCB1	**91.14%**	**158.3**	**91.01%**	**0.037**	87.54%	0.168	**88.35%**	1256.2
*UCB1-R	**91.12%**	**164.3**	**89.76%**	**0.035**	86.28%	0.169	86.48%	1179.6
*TS	86.86%	149.1	**88.84%**	**0.035**	69.93%	0.162	85.41%	1155.1
*TS-R	**88.72%**	**156.3**	**88.94%**	**0.035**	73.09%	0.164	85.64%	1164.8
	ETTm1		ETTm2		ETTh1		ETTh2	
Method	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width
CQR	77.33%	6.386	92.88%	14.670	75.95%	7.858	93.79%	14.127
ACI	**89.93%**	13.050	**88.49%**	21.383	**90.52%**	27.618	**90.34%**	31.509
Asym-ACI	**90.75%**	**7.587**	**88.08%**	**14.165**	92.65%	14.046	**90.94%**	**18.201**
*UCB1	**91.76%**	**7.851**	**89.30%**	**15.067**	**91.97%**	**13.963**	91.03%	**19.107**
*UCB1-R	**89.63%**	**7.942**	**89.25%**	**15.598**	92.50%	14.431	**90.69%**	**19.164**
*TS	87.10%	7.544	83.98%	14.356	86.97%	13.462	**88.17%**	**18.386**
*TS-R	**88.81%**	**7.741**	87.20%	15.054	**88.20%**	**13.665**	**88.75%**	**18.552**

Table 3. General Dataset Experiment

6.6 Conclusion

Empirically speaking, ASACI can produce a valid conformal set in most asymmetric cases. In general cases, ASACI can also attain a targeted level with a smaller interval. As observed from the experience, ASACI with the UCB1 bandit is often slightly higher than the targeted conformal level, while Thompson-Sampling is often slightly lower. Thus, a general rule of thumb for choosing between UCB1 and Thompson-Sampling is that for a harder task or scenario where more conservatism is needed, UCB1 will be a good choice. Otherwise, Thompson-Sampling can be applied in cases where a small interval is desired.

Acknowledgement

We thank the anonymous reviewers and members of CLLab for their constructive feedback. This work is partially supported by the National Science and Technology Council in Taiwan via NSTC 113-2634-F-002-008, 114-2221-E-002-102-MY3, and the National Taiwan University Center for Data Intelligence via NTU-114L900901. We thank to National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) in Taiwan for providing computational and storage resources. H.-T. Lin is honored to be supported by the Leap Fellowship of the Foundation for the Advancement of Outstanding Scholarship in Taiwan since 2025.

References

1. Abad, J., Bhatt, U., Weller, A., Cherubin, G.: Approximating full conformal prediction at scale via influence functions (2023)
2. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion **76**, 243–297 (2021)

3. Angelopoulos, A., Bates, S., Malik, J., Jordan, M.I.: Uncertainty sets for image classifiers using conformal prediction (2022)
4. Angelopoulos, A.N., Bates, S.: A gentle introduction to conformal prediction and distribution-free uncertainty quantification (2022)
5. Angelopoulos, A.N., Candes, E.J., Tibshirani, R.J.: Conformal pid control for time series prediction (2023)
6. Auer, A., Gauch, M., Klotz, D., Hochreiter, S.: Conformal prediction for time series with modern hopfield networks (2023)
7. Barber, R.F., Candes, E.J., Ramdas, A., Tibshirani, R.J.: Predictive inference with the jackknife+ (2020)
8. Barber, R.F., Candes, E.J., Ramdas, A., Tibshirani, R.J.: Conformal prediction beyond exchangeability (2023)
9. Bhatnagar, A., Wang, H., Xiong, C., Bai, Y.: Improved online conformal prediction via strongly adaptive online learning (2023)
10. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics **31**(3), 307–327 (1986)
11. Chen, S.: Beijing PM2.5. UCI Machine Learning Repository (2015)
12. Christoffersen, P.F., Diebold, F.X.: Optimal prediction under asymmetric loss. Econometric Theory **13**(6), 808–817 (1997)
13. Foygel Barber, R., Candès, E.J., Ramdas, A., Tibshirani, R.J.: The limits of distribution-free conditional predictive inference. Information and Inference: A Journal of the IMA **10**(2), 455–482 (2020)
14. Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R., Zhu, X.X.: A survey of uncertainty in deep neural networks (2022)
15. Gibbs, I., Candes, E.: Adaptive conformal inference under distribution shift. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems. vol. 34 (2021)
16. Gibbs, I., Candès, E.: Conformal inference for online prediction with arbitrary distribution shifts (2023)
17. He, W., Jiang, Z.: A survey on uncertainty quantification methods for deep neural networks: An uncertainty source perspective (2023)
18. Jensen, V., Bianchi, F.M., Anfinsen, S.N.: Ensemble conformalized quantile regression for probabilistic time series forecasting. IEEE Transactions on Neural Networks and Learning Systems p. 1–12 (2022)
19. Lai, G., Chang, W.C., Yang, Y., Liu, H.: Modeling long- and short-term temporal patterns with deep neural networks. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. p. 95–104. SIGIR '18, Association for Computing Machinery, New York, NY, USA (2018)
20. Lei, J., Wasserman, L.: Distribution-free Prediction Bands for Non-parametric Regression. Journal of the Royal Statistical Society Series B: Statistical Methodology **76**(1), 71–96 (07 2013)
21. Ndiaye, E., Takeuchi, I.: Computing full conformal prediction set with approximate homotopy (2019)
22. Ndiaye, E., Takeuchi, I.: Root-finding approaches for computing conformal prediction set. Machine Learning **112**(1) (2022)
23. Oliveira, R.I., Orenstein, P., Ramos, T., Romano, J.V.: Split conformal prediction for dependent data (2022)
24. Papadopoulos, H., Proedrou, K., Vovk, V., Gammerman, A.: Inductive confidence machines for regression. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) Machine Learning: ECML 2002 (2002)

25. Park, S., Dobriban, E., Lee, I., Bastani, O.: Pac prediction sets under covariate shift (2022)
26. Podkopaev, A., Ramdas, A.: Distribution-free uncertainty quantification for classification under label shift (2021)
27. Qiu, H., Dobriban, E., Tchetgen, E.T.: Prediction sets adaptive to unknown covariate shift (2023)
28. Romano, Y., Patterson, E., Candes, E.: Conformalized quantile regression. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 32 (2019)
29. Romano, Y., Sesia, M., Candès, E.J.: Classification with valid and adaptive coverage (2020)
30. Shafer, G., Vovk, V.: A tutorial on conformal prediction (2007)
31. Sun, S., Yu, R.: Copula conformal prediction for multi-step time series forecasting (2024)
32. Tibshirani, R.J., Barber, R.F., Candes, E.J., Ramdas, A.: Conformal prediction under covariate shift (2020)
33. Trindade, A.: ElectricityLoadDiagrams20112014. UCI Machine Learning Repository (2015)
34. Vovk, V.: Cross-conformal predictors (2012)
35. Vovk, V., Gammerman, A., Shafer, G.: Algorithmic Learning in a Random World. Springer-Verlag, Berlin, Heidelberg (2005)
36. Vovk, V., Nouretdinov, I., Manokhin, V., Gammerman, A.: Cross-conformal predictive distributions. In: Gammerman, A., Vovk, V., Luo, Z., Smirnov, E., Peeters, R. (eds.) Proceedings of the Seventh Workshop on Conformal and Probabilistic Prediction and Applications. Proceedings of Machine Learning Research, vol. 91, pp. 37–51. PMLR (11–13 Jun 2018)
37. Vovk, V., Gammerman, A., Saunders, C.: Machine-learning applications of algorithmic randomness. In: Proceedings of the Sixteenth International Conference on Machine Learning. p. 444–453. ICML '99, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1999)
38. Wu, J., Hu, D., Bao, Y., Xia, S.T., Zou, C.: Error-quantified conformal inference for time series. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=RD9q5vEe1Q
39. Xu, C., Xie, Y.: Conformal prediction interval for dynamic time-series. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 11559–11569. PMLR (2021)
40. Xu, C., Xie, Y.: Conformal prediction set for time-series (2022)
41. Xu, C., Xie, Y.: Conformal prediction for time series (2023)
42. Xu, C., Xie, Y.: Sequential predictive conformal inference for time series (2023)
43. Yang, Y., Kuchibhotla, A.K., Tchetgen, E.T.: Doubly robust calibration of prediction sets under covariate shift (2022)
44. Zaffran, M., Dieuleveut, A., Féron, O., Goude, Y., Josse, J.: Adaptive conformal predictions for time series (2022)
45. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? (2023)
46. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference (2021)