

# A Review and Comparative Analysis of Univariate Conformal Regression Methods

**Jie Bao**

*Huaiyin Institute of Technology, Huai'an, China*

1486103897@QQ.COM

**Nicolo Colombo**

*Royal Holloway, University of London, Egham, Surrey, UK*

NICOLO.COLOMBO@RHUL.AC.UK

**Valery Manokhin**

*Independent Researcher*

VALERY.MANOKHIN.2015@LIVE.RHUL.AC.UK

**Suqun Cao**

*Huaiyin Institute of Technology, Huai'an, China*

CAOSUQUN@HYIT.EDU.CN

**Rui Luo**

*City University of Hong Kong, Hong Kong, China*

RUILUO@CITYU.EDU.HK

**Editor:** Khuong An Nguyen, Zhiyuan Luo, Harris Papadopoulos, Tuwe Löfström, Lars Carlsson and Henrik Boström

## Abstract

As machine learning models continue to evolve and improve, quantifying their uncertainty has become increasingly crucial in high-stakes applications. Conformal prediction has emerged as a powerful tool and has been widely applied in univariate regression tasks. While numerous conformal regression methods and models have been developed, few studies have provided a unified summary and comparison of these approaches. In this paper, we address this gap by discussing, summarizing, and providing an overview of the majority of existing univariate conformal regression methods. Furthermore, we conduct a detailed examination and experimentation of eight major, popular, and advanced conformal regression methods, representing a significant contribution to the field by offering a comprehensive analysis and insights into their performance and applicability.

**Keywords:** conformal prediction, exchangeable, conformal regression, uncertainty quantification.

## 1. Introduction

In many real-world applications, accurately quantifying the uncertainty of model predictions is crucial for risk management, especially in high-stakes fields such as medical diagnosis (Vazquez and Facelli, 2022; Luo et al., 2024), autonomous driving (Lindemann et al., 2023; Chen et al., 2024), and financial risk control (Angelopoulos et al., 2022; Overman et al., 2024). Conformal Prediction (CP) (Fontana et al., 2023; Angelopoulos et al., 2023) presents a statistically rigorous framework that provides probabilistic guarantees, transforming point predictions from traditional machine learning models into prediction intervals with well-defined coverage probabilities. This framework, grounded in the fundamental assumption of data exchangeability, is capable of generating prediction intervals that adhere to a pre-established confidence level without any reliance on the particular model architecture. In recent years, CP (Huang et al., 2024) has exhibited notable efficacy across various machine learning realms, encompassing regression analysis (Romano et al., 2019), multi-class decision-making (Luo and Zhou, 2024), and beyond.

Conformal prediction, as a method for generating valid confidence intervals without assuming the data-generating distribution  $P(X|Y)$  or the prediction model  $f$  (Vovk et al., 2005; Shafer and Vovk, 2008), and has gained increasing attention. Currently, univariate conformal regression is the primary research direction in this field. Many scholars have developed extensions of this approach, including Conformal Quantile Regression (Romano et al., 2019), Conformal Histogram Regression

(Sesia and Romano, 2021), Conformal Thresholded Intervals (Luo and Zhou, 2025b), etc., all of which have made significant contributions to the advancement of conformal regression.

In this paper, we provide a comprehensive survey and comparative analysis of the state-of-the-art conformal regression methods. Our main contributions are as follows:

- We offer a detailed categorization and discussion of a wide range of existing conformal regression techniques.
- We examine and summarize the underlying principles and models of eight leading and cutting-edge conformal regression methods.
- We visualize the differences in prediction intervals produced by these methods through simulation experiments.
- We perform experiments on twelve real-world datasets using these eight methods, followed by an in-depth comparison and discussion of the results.

To present these contributions in a clear and structured manner, the remainder of the paper is organized as follows. Section 2 establishes the necessary background on conformal prediction, reviewing its foundational concepts and recent methodological advancements. Section 3 provides a detailed exposition of the eight distinct conformal regression methods that are the focus of our study, elaborating on their underlying mechanics and theoretical differences. Section 4 is dedicated to our extensive empirical evaluation, beginning with a simulation study to visually illustrate the behavior of each method, followed by a rigorous quantitative comparison on twelve real-world datasets. Finally, Section 5 concludes the paper by summarizing our key findings, discussing the relative strengths and weaknesses of the evaluated approaches, and outlining promising directions for future research.

## 2. Background

Conformal prediction (CP) Vovk et al. (2005) is a methodology designed to generate prediction regions for variables of interest, facilitating the estimation of model uncertainty by providing prediction sets rather than point estimates. CP has been successfully applied to both classification Luo and Zhou (2024); Luo and Colombo (2024); Luo and Zhou (2025d) and regression tasks Luo and Zhou (2025e,f). Its flexibility allows adaptation to various real-world scenarios, including segmentation Luo and Zhou (2025a), games Luo et al. (2024); Bao et al. (2025), time-series forecasting Su et al. (2024), and graph-based applications Luo et al. (2023); Tang et al. (2025); Luo and Zhou (2025c); Wang et al. (2025); Luo and Zhou (2025b); Zhang et al. (2025).

Consider a univariate regression problem where the objective is to predict a scalar response  $y \in \mathcal{Y} = \mathbb{R}$  based on a feature vector  $x \in \mathcal{X} \subseteq \mathbb{R}^p$ . We assume that there exists a true joint distribution  $F_{XY}$  over  $\mathcal{X} \times \mathcal{Y}$ , and we have access to a dataset

$$\mathcal{D} = \{(x_{(i)}, y_{(i)})\}_{i=1}^n, \quad (1)$$

where the pairs  $(x_{(i)}, y_{(i)})$  are independent and identically distributed according to  $F_{XY}$ . Given a new feature vector  $x$ , we denote the conditional distribution of  $Y$  given  $X = x$  as  $F_{Y|X=x}$  and the corresponding probability density function as  $f_{Y|X=x}$  (when it exists). Using the dataset  $\mathcal{D}$ , CP allow us to transform any point predictor, denoted by  $\hat{h}$ , into a calibrated, distribution-free prediction interval  $\hat{C}(x) \subseteq \mathcal{Y}$  for the true response  $y$ . These intervals come with finite-sample coverage guarantees of the form:

$$P(Y \in \hat{C}(X)) \geq 1 - \alpha, \quad (2)$$

where  $1 - \alpha$  is the desired coverage level.

## 2.1. Split Conformal Prediction

Traditional split conformal prediction methods ([Vovk et al., 1999](#); [Papadopoulos et al., 2002](#); [Vovk et al., 2005, 2009](#); [Lei et al., 2013](#)) guarantee marginal coverage in finite samples by randomly partitioning data into training and calibration sets, then constructing prediction intervals for new samples using quantiles of nonconformity scores computed on the calibration data. However, these methods often suffer from the following limitations:

- **Fixed Width:** Standard split CP constructions typically rely on global score distributions, which results in prediction intervals with nearly uniform width across all inputs, failing to adequately adapt to local heterogeneity in conditional distributions.
- **Simple Residual Calibration:** Most approaches use absolute residuals as nonconformity scores, without fully leveraging local data information, potentially yielding prediction intervals that are either excessively conservative or insufficiently precise in certain regions.

## 2.2. Advances in Base Regression Models for CP

The performance of the base regression model has a decisive impact on the quality of prediction intervals within the conformal regression framework ([Magdon-Ismail and Atiya, 1998](#); [Meinshausen and Ridgeway, 2006](#); [Chipman et al., 2010](#); [Kivaranovic et al., 2020](#); [Moon et al., 2021](#); [Du et al., 2022](#)). Researchers have significantly enhanced the effectiveness of CP by integrating advanced regression techniques across different methodological paradigms:

Early studies focused on combining CP with classical regression models. For example, [Lei and Wasserman \(2014\)](#) proposed using linear regression to model local patterns, while [Johansson et al. \(2014b\)](#) introduced decision tree-based approaches to adaptively partition the feature space. These foundational works demonstrated the potential of leveraging model-specific structures to improve interval adaptivity.

Subsequent research expanded into more sophisticated techniques. [Papadopoulos et al. \(2008, 2011\)](#) pioneered localized modeling by constructing prediction intervals using k-nearest neighbor regression, directly incorporating neighborhood information into nonconformity scores. [Johansson et al. \(2014a\)](#) combined random forests with quantile regression, thereby leveraging ensemble learning to enhance model robustness and handle heterogeneous data distributions. Furthermore, in the domain of deep learning, [Johansson et al. \(2015\)](#) developed a bagged neural network quantile regression model, while [Romano et al. \(2019\)](#) proposed a two-model architecture using separate quantile regressors for interval bounds.

Recent innovations further address computational and theoretical challenges. [Kivaranovic et al. \(2020\)](#) introduced a framework that divides the prediction into three parts: lower bounds, medians, and upper bounds to capture complex data patterns. [Boström et al. \(2017\)](#) optimized quantile estimation processes for random forests through acceleration algorithms. Notably, [Gibbs et al. \(2025\)](#) embedded test point features directly into an augmented quantile regression model to explicitly control conditional coverage. Extensions like [Sousa et al. \(2024\)](#) handle heteroscedasticity via dynamic variance estimation, while [Rosenberg et al. \(2022\)](#) unified multivariate predictions through nonlinear vector quantile regression.

Collectively, these advances aim to reduce model uncertainty while maintaining the theoretical rigor of conformal methods, ensuring precise marginal coverage and progress toward conditional coverage guarantees.

### 2.3. Recent Improvements in Prediction Interval Construction

Building upon the foundational frameworks of conformal prediction, recent methodological developments have introduced significant refinements in constructing adaptive prediction intervals. These improvements primarily focus on three key technical directions: conditional density estimation, residual-based calibration, and localized adaptation strategies, each offering unique advantages in constructing adaptive prediction intervals.

**Conditional Density Estimation Approaches.** A key direction involves leveraging explicit estimates of conditional distributions  $f_{Y|X=x}$  to construct locally adaptive intervals. For instance, Izbicki et al. (2019) utilize conditional density functions to model data distributions directly, enabling theoretically optimal interval construction under distributional assumptions. The SPICE framework (Diamant et al., 2024) employs neural networks to estimate  $f_{Y|X=x}$ , while Plassier et al. (2024) propose the CP2 method combining conformity score transformations with conditional density estimation to achieve approximate conditional validity.

**Residual Distribution Methods.** Alternative approaches exploit residual distributions to adjust interval widths. Chen et al. (2018) introduced a method for determining prediction intervals by analyzing residual distributions and considering specified significance levels. Building on similar ideas, Barber et al. (2021) proposed constructing prediction intervals centered around the median (or alternatively, the mean) of predictions obtained through a leave-one-out approach. Meanwhile, Lei et al. (2018) developed the Split Conformal Prediction Sets method, which involves randomly partitioning the data into two subsets,  $I_1$  and  $I_2$ . The model  $\mu$  is trained on  $I_1$ , and residuals are computed and sorted on  $I_2$ . Prediction intervals are then constructed based on these sorted residuals and a pre-selected  $\alpha$  level, ensuring a desired level of confidence. More recently, Luo and Zhou (2025c) refined this approach by calculating residuals between predicted and true labels within each subset and leveraging these residuals to construct more robust prediction sets.

**Localized Adaptation Strategies.** To achieve conditional coverage guarantees, some methods employ spatial adaptation mechanisms. The RLCP method (Hore and Barber, 2025) introduces randomized local weighting to adjust conformity thresholds, while Kiyani et al. (2024a) (PLCP) partitions the covariate space into regions with homogeneous uncertainty levels. Colombo (2024) propose redefining conformity measures  $\phi_x(\cdot)$  to explicitly depend on  $X$ , enabling input-dependent interval scaling. For covariate shift scenarios, Wieczorek (2023) adjust score distributions through sampling weights derived from design-based estimators. Complementary approaches proposed by van der Laan and Alaa (2024) generate instance-specific calibration sets, and Gil et al. (2024) identify homogeneous regions where conformity scores exhibit uniform distributions. Additionally, Cheung et al. (2024) address interval asymmetry via bias-aware adjustments.

### 2.4. Other Works in Conformal Regression

Existing works extend conformal regression along two main directions: tight integration with regression training objectives and adaptations to real-world regression challenges.

**Training-aware methods.** Recent advances embed conformal principles directly into model training processes. Three primary approaches emerge: (a) Bilevel optimization frameworks that co-optimize regression parameters and prediction interval thresholds (Kiyani et al., 2024b); (b) Hybrid loss functions unifying point estimation and uncertainty calibration through terms like distributional alignment (Pouplin et al., 2024) and coverage-aware regularization (Gao et al., 2024); (c) Adaptive mechanisms including input-dependent prediction distributions (Vovk et al., 2020) and spatially-aware density estimators (Plassier et al., 2024), which dynamically adjust intervals using

local uncertainty patterns. There is also a part of work [Bellotti \(2021\)](#); [Stutz et al. \(2022\)](#); [Colombo \(2023\)](#) that proposes training conformal prediction on classification tasks to achieve smaller uncertainty.

**Real-world adaptations.** Methodological extensions address practical constraints: Cross-validation conformal methods ([Vovk, 2015](#)) improve computational scalability, while specialized variants handle federated learning ([Humbert et al., 2023](#)), missing covariates ([Zaffran et al., 2023](#)), and discrete responses ([Sesia et al., 2023](#)). Spatial calibration techniques ([Meister and Nguyen, 2025](#)) maintain validity under distribution shifts through auxiliary variable encoding and geometric sketching. These innovations preserve conformal methods' finite-sample coverage guarantees while expanding applicability to complex data ecosystems.

### 3. Conformal Prediction for Regression

In this section, we will provide a detailed introduction to several distinct methods of conformal prediction for regression that have been developed to date. We will elaborate on their connections and differences in sequence. These methods represent the most popular and cutting-edge approaches currently available.

**Split Conformal Regression** ([Papadopoulos et al., 2002](#); [Vovk et al., 2005](#)). First, we need to randomly partition a given dataset into a training set, a calibration set, and a test set, denoted as  $\mathcal{I}_{\text{train}}$ ,  $\mathcal{I}_{\text{cal}}$ , and  $\mathcal{I}_{\text{test}}$ , respectively. We require a basic predictive model  $\hat{y}_i = \hat{f}(x_i)$ , which is trained on the training set  $\mathcal{I}_{\text{train}}$ . Subsequently, we compute the conformity scores  $S_i$ , on the calibration set  $\mathcal{I}_{\text{cal}}$ :

$$S_i^{\text{split}} = |y_i - \hat{f}(x_i)|. \quad (3)$$

*Split conformal regression* utilizes the absolute residuals to determine the  $1 - \alpha$  prediction interval for a new test point  $x_{n+1} \in \mathcal{I}_{\text{test}}$ :

$$\mathcal{C}_{n,\alpha}(x_{n+1}) = \{y \in \mathbb{R} : s(x_{n+1}, y) \leq t_{1-\alpha}^{\text{split}}\}, \quad (4)$$

where  $t_{1-\alpha}^{\text{split}}$  is the  $(1 - \alpha)(1 + 1/|\mathcal{I}_{\text{cal}}|)$ -th empirical quantile of  $\{S_i^{\text{split}}\}_{i \in \mathcal{I}_{\text{cal}}} \cup \{\infty\}$ . Under the exchangeability assumption, this guarantees marginal coverage:

$$\mathbb{P}(Y_{n+1} \in \mathcal{C}(x_{n+1})) \geq 1 - \alpha. \quad (5)$$

**Conformal Quantile Regression (1)** ([Romano et al., 2019](#)). *Conformal Quantile Regression* constructs intervals based on quantile regression:

$$\mathcal{C}^{\text{CQR}}(x_{n+1}) = \left[ \hat{q}_{\frac{\alpha}{2}}(x_{n+1}) - t_{1-\alpha}^{\text{CQR}}, \hat{q}_{1-\frac{\alpha}{2}}(x_{n+1}) + t_{1-\alpha}^{\text{CQR}} \right], \quad (6)$$

where  $\hat{q}_{\alpha/2}$  and  $\hat{q}_{1-\alpha/2}$  are conditional quantile estimates, and  $t_{1-\alpha}^{\text{CQR}}$  is the  $(1 - \alpha)(1 + 1/|\mathcal{I}_{\text{cal}}|)$ -th empirical quantile of  $\{S_i^{\text{CQR}}\}_{i \in \mathcal{I}_{\text{cal}}} \cup \{\infty\}$ , with:

$$S_i^{\text{CQR}} = \max \left( \hat{q}_{\frac{\alpha}{2}}(x_i) - y_i, y_i - \hat{q}_{1-\frac{\alpha}{2}}(x_i) \right). \quad (7)$$

**Conformal Quantile Regression (2) (Kivaranic et al., 2020).** Similar to Sesia and Candès (2020), we name the methods proposed by Kivaranic et al. (2020) as CQR-m, respectively, which differ from the CQR proposed by Romano et al. (2019). First, let's consider CQR-m. The model defined based on quantile regression is as follows:

$$\begin{aligned}\mathcal{C}^{\text{CQR-m}}(x_{n+1}) &= \left[ \hat{q}_{\frac{\alpha}{2}}(x_{n+1}) - \hat{\Delta}_{\alpha,\text{lo}}^{\text{CQR-m}}, \hat{q}_{1-\frac{\alpha}{2}}(x_{n+1}) + \hat{\Delta}_{\alpha,\text{up}}^{\text{CQR-m}} \right], \\ \hat{\Delta}_{\alpha,\text{lo}}^{\text{CQR-m}} &= t_{1-\alpha}^{\text{CQR-m}} \left[ \hat{q}_{\frac{1}{2}}(x_i) - \hat{q}_{\frac{\alpha}{2}}(x_i) \right], \\ \hat{\Delta}_{\alpha,\text{up}}^{\text{CQR-m}} &= t_{1-\alpha}^{\text{CQR-m}} \left[ \hat{q}_{1-\frac{\alpha}{2}}(x_i) - \hat{q}_{\frac{1}{2}}(x_i) \right],\end{aligned}\quad (8)$$

where  $\hat{q}_{1/2}$  indicates an estimated median regression function obtained with the same black-box algorithm as  $\hat{q}_{\alpha/2}$  and  $\hat{q}_{1-\alpha/2}$ , and  $t_{1-\alpha}^{\text{CQR-m}}$  is the same  $(1-\alpha)(1+1/|\mathcal{I}_{\text{cal}}|)$ -th empirical quantile of  $\{S_i^{\text{CQR-m}}\}_{i \in \mathcal{I}_{\text{cal}}} \cup \{\infty\}$ , with:

$$S_i^{\text{CQR-m}} = \max \left( \frac{\hat{q}_{\frac{\alpha}{2}}(x_i) - y_i}{\hat{q}_{\frac{1}{2}}(x_i) - \hat{q}_{\frac{\alpha}{2}}(x_i)}, \frac{y_i - \hat{q}_{1-\frac{\alpha}{2}}(x_i)}{\hat{q}_{1-\frac{\alpha}{2}}(x_i) - \hat{q}_{\frac{1}{2}}(x_i)} \right). \quad (9)$$

In addition, there is an improved version of CQR-m, which does not require estimating the quantile of the regression median (Sesia and Candès, 2020). The prediction interval it constructs is as follows:

$$\begin{aligned}\mathcal{C}^{\text{CQR-r}}(x_{n+1}) &= \left[ \hat{q}_{\frac{\alpha}{2}}(x_{n+1}) - \hat{\Delta}_{\alpha}^{\text{CQR-r}}, \hat{q}_{1-\frac{\alpha}{2}}(x_{n+1}) + \hat{\Delta}_{\alpha}^{\text{CQR-r}} \right], \\ \hat{\Delta}_{\alpha}^{\text{CQR-r}} &= t_{1-\alpha}^{\text{CQR-r}} \left[ \hat{q}_{1-\frac{\alpha}{2}}(x_i) - \hat{q}_{\frac{\alpha}{2}}(x_i) \right],\end{aligned}\quad (10)$$

where  $t_{1-\alpha}^{\text{CQR-r}}$  is the  $(1-\alpha)(1+1/|\mathcal{I}_{\text{cal}}|)$ -th empirical quantile of  $\{S_i^{\text{CQR-r}}\}_{i \in \mathcal{I}_{\text{cal}}} \cup \{\infty\}$ , with:

$$S_i^{\text{CQR-r}} = \max \left( \frac{\hat{q}_{\frac{\alpha}{2}}(x_i) - y_i}{\hat{q}_{1-\frac{\alpha}{2}}(x_i) - \hat{q}_{\frac{\alpha}{2}}(x_i)}, \frac{y_i - \hat{q}_{1-\frac{\alpha}{2}}(x_i)}{\hat{q}_{1-\frac{\alpha}{2}}(x_i) - \hat{q}_{\frac{1}{2}}(x_i)} \right). \quad (11)$$

**Conformal Quantile Regression with Full Model (Kivaranic et al., 2020).** CQRFM builds upon CQR-m by introducing a modification that allows the model to output three distinct values: the lower bound, median, and upper bound simultaneously from a single neural network. The key idea is to train a neural network  $\mathcal{N} : \mathbb{R}^d \rightarrow \mathbb{R}^3$  such that  $\mathcal{N}(x) = (l(x), m(x), u(x))$ , where  $l$ ,  $m$ , and  $u$  are functions that estimate the  $\alpha/2$ -quantile, the median, and the  $(1-\alpha/2)$ -quantile, respectively, with the constraint that  $l(x) \leq m(x) \leq u(x)$  for all  $x \in \mathbb{R}^d$ .

The network is trained using a modified quantile regression loss function:

$$L_\tau(\mathcal{N}(x), y) = h_{\tau/2}(y - l(x)) + h_{1/2}(y - m(x)) + h_{1-\tau/2}(y - u(x)), \quad (12)$$

where  $h_\tau(u) = (\tau - \mathbf{1}_{u \leq 0})u$  is the standard quantile regression loss function.

Similar to CQR-m, the prediction interval is constructed as:

$$\begin{aligned}\mathcal{C}^{\text{CQRFM}}(x_{n+1}) &= \left[ l(x_{n+1}) - \hat{\Delta}_{\alpha,\text{lo}}^{\text{CQRFM}}, u(x_{n+1}) + \hat{\Delta}_{\alpha,\text{up}}^{\text{CQRFM}} \right], \\ \hat{\Delta}_{\alpha,\text{lo}}^{\text{CQRFM}} &= t_{1-\alpha}^{\text{CQRFM}} [m(x_i) - l(x_i)], \\ \hat{\Delta}_{\alpha,\text{up}}^{\text{CQRFM}} &= t_{1-\alpha}^{\text{CQRFM}} [u(x_i) - m(x_i)],\end{aligned}\quad (13)$$

where  $t_{1-\alpha}^{\text{CQRFM}}$  is the  $(1-\alpha)(1+1/|\mathcal{I}_{\text{cal}}|)$ -th empirical quantile of  $\{S_i^{\text{CQRFM}}\}_{i \in \mathcal{I}_{\text{cal}}} \cup \{\infty\}$ , with:

$$S_i^{\text{CQRFM}} = \max \left( \frac{l(x_i) - y_i}{m(x_i) - l(x_i)}, \frac{y_i - u(x_i)}{u(x_i) - m(x_i)} \right). \quad (14)$$

**Conformal Histogram Regression (Sesia and Romano, 2021).** *Conformal Histogram Regression* constructs prediction intervals by estimating the full conditional density  $f_{Y|X}$  using histograms and finding the shortest interval  $(a, b)$  such that:

$$\mathcal{C}^{\text{CHR}}(x_{n+1}) = \arg \min_{a < b} (b - a), \quad (15)$$

$$\text{s.t. } \int_a^b \hat{f}_{Y|X}(y|x_{n+1}) dy \geq 1 - \alpha. \quad (16)$$

**Localized Conformal Prediction (Guan, 2023).** *Localized Conformal Prediction* (LCP) generalizes the framework of conformal prediction by offering a single-test-sample adaptive construction that emphasizes a local region around the test sample. LCP introduces a localizer function  $H(x, x') : \mathbb{R}^p \times \mathbb{R}^p \rightarrow [0, 1]$  that captures the similarity between feature values, with  $H(x, x) = 1$  for all  $x$ . For a test point  $x_{n+1}$ , LCP assigns different weights to calibration samples based on their proximity to  $x_{n+1}$ .

Let  $H_{n+1,i} = H(x_{n+1}, x_i)$  be the localizer evaluated at  $x_{n+1}$  and  $x_i$ , and define the weighted empirical distribution:

$$\hat{\mathcal{F}}_{n+1} = \sum_{j=1}^n p_{n+1,j}^H \delta_{S_j} + p_{n+1,n+1}^H \delta_\infty, \quad (17)$$

where  $p_{i,j}^H = H_{ij} / \sum_{k=1}^{n+1} H_{ik}$  for  $j = 1, \dots, n+1$  are the normalized weights. To ensure finite-sample marginal coverage, LCP requires finding a suitable  $\alpha$  that satisfies:

$$(n+1)^{-1} \sum_{i=1}^{n+1} \mathbb{1}_{S_i \leq Q(\tilde{\alpha}; \hat{\mathcal{F}}_i)} \geq 1 - \alpha, \quad (18)$$

where  $\hat{\mathcal{F}}_i$  is the weighted empirical distribution with weights centered at  $x_i$ , and  $Q(\tilde{\alpha}; \hat{\mathcal{F}}_i)$  is the  $\tilde{\alpha}$ -quantile of this distribution. The prediction interval for the test point  $x_{n+1}$  is then given by:

$$\mathcal{C}^{\text{LCP}}(x_{n+1}) = \left\{ y \in \mathbb{R} : s(x_{n+1}, y) \leq Q(\tilde{\alpha}^*; \hat{\mathcal{F}}_{n+1}) \right\} \quad (19)$$

LCP can be versatilely paired with a diverse range of conformal scores. For instance, it utilizes the standard residual score (3).

**Conformal Thresholded Intervals (Luo and Zhou, 2025b).** *Conformal Thresholded Intervals* achieves prediction set size minimization through selective inclusion of narrower probability ranges. This mechanism operationalizes the confidence set definition as:

$$\mathcal{C}^{\text{CTI}}(x_{n+1}) == \bigcup \{ I_k(x_t) : \mu(I_k(x_t)) \leq S_i^{\text{CTI}}, k = 1, \dots, K \}. \quad (20)$$

where  $S_i^{\text{CTI}}$  is the  $(1 - \alpha)$ -th quantile of the empirical distribution:

$$S_i^{\text{CTI}} = \frac{1}{(1 + |I_{\text{cal}}|)} \sum_{i \in I_{\text{cal}}} \delta_{\mu(I_{k(y_i)}(x_i))} + \delta_\infty. \quad (21)$$

where  $k$  is the given quantile step size, and  $I_k(x_i) = (\hat{q}_{k-1}(x_i), \hat{q}_k(x_i)]$  for  $k = 1, \dots, K$ .

We conducted a comparative analysis of eight distinct conformal regression methods (1). Our findings indicate that CTI and CHR generate narrower prediction intervals while demonstrating superior robustness in handling data with conditional distribution shifts. We theoretically analyzed

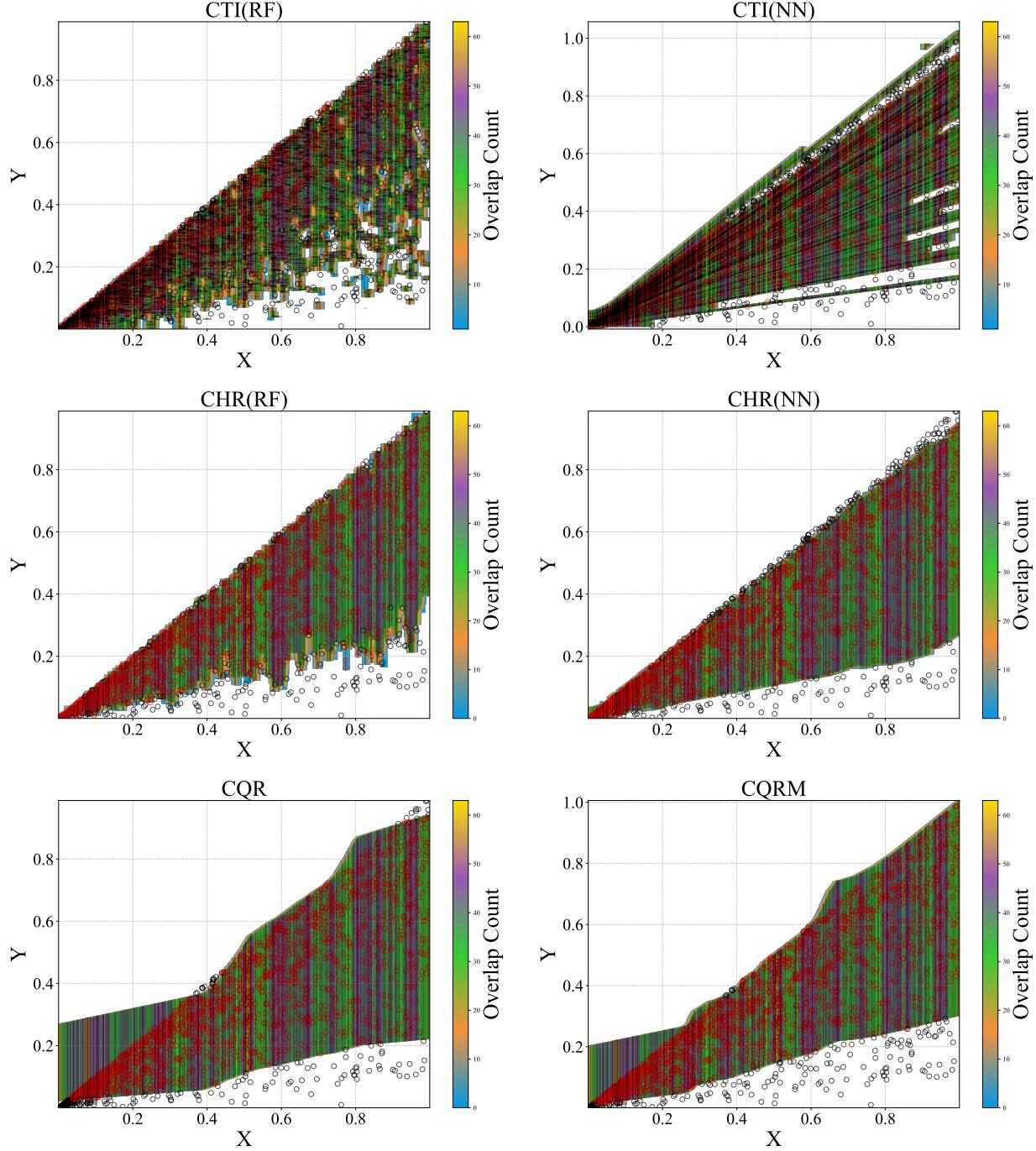


Figure 1: In the figure, the scatter points with coordinates  $(x, y)$  represent features  $(x)$  and true labels  $(y)$ . Any conformal regression method will generate a prediction interval for each of these scatter points. For the purpose of visualizing the prediction intervals produced by each method, we set a uniform width of 0.01 for these intervals, while the height of each interval is determined by the conformal regression model. This approach ensures that the rectangular boxes representing each model's intervals have consistent width but varying heights. Naturally, wider prediction intervals result in taller rectangular boxes, and overlapping may occur between adjacent boxes. The denser the overlapping regions, the larger the size of the resulting prediction set. To facilitate visualization, we use different colors to represent the density of these generated intervals, allowing us to intuitively discern the differences between these methods. Additionally, we depict scatter points that fall within their respective prediction intervals in red, while those outside the intervals are shown in black. The same visualization approach is applied to Figure 2.

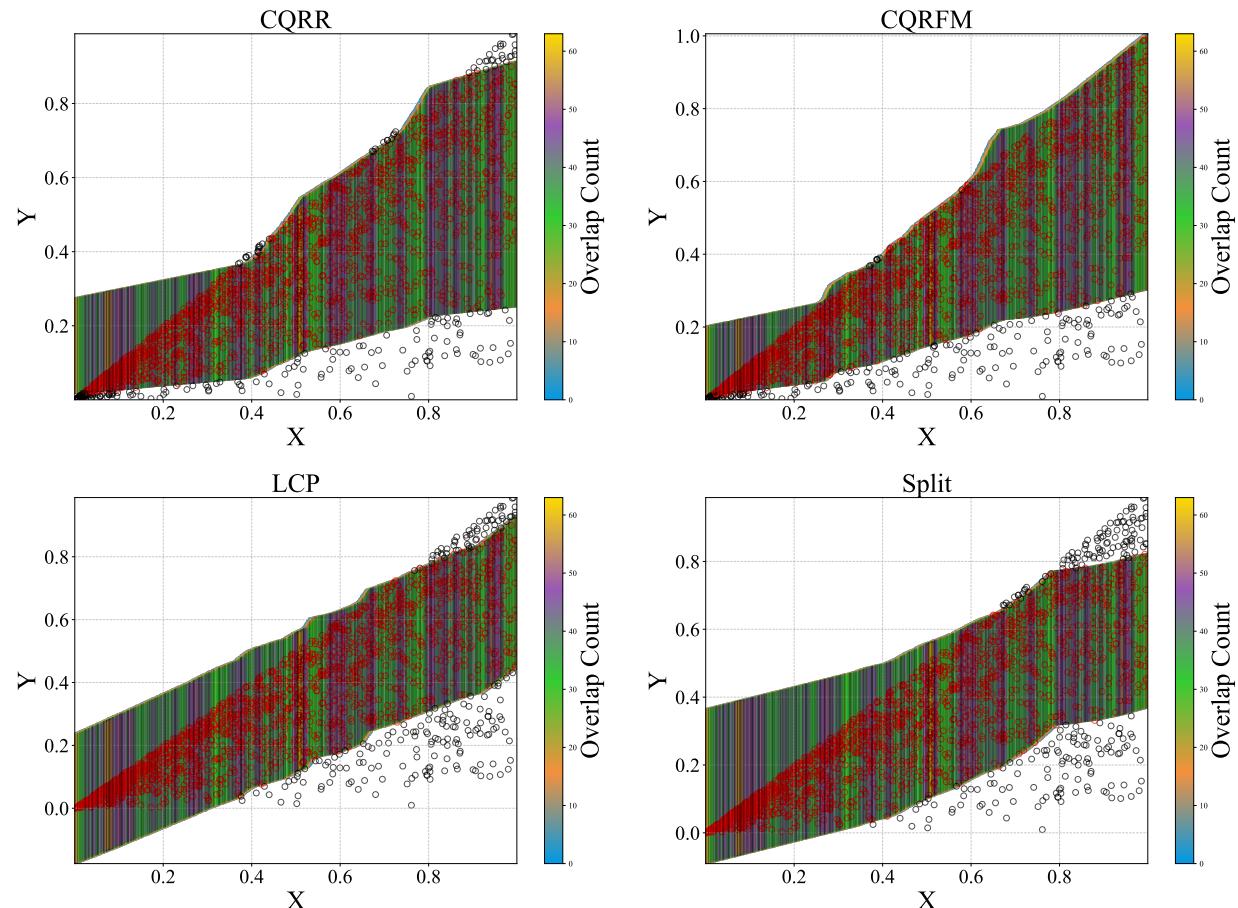


Figure 2: Synthetic figures.

Method	basic model	Additional resource investment	Robustness to Conditional Shift	Small average size	Asymptotic time complexity	Asymptotic space complexity	Predictive interval continuous
Split	Point prediction model	✗	★★☆☆☆	★☆☆☆☆	$O(N \log N)$	$O(N)$	✓
CQR	Quantile regression model	✓	★★☆☆☆	★★☆☆☆	$O(N \log N)$	$O(N)$	✓
CQR-m	Quantile regression model	✗	★★☆☆☆	★★☆☆☆	$O(N \log N)$	$O(N)$	✓
CQR-r	Quantile regression model	✗	★☆☆☆☆	★★☆☆☆	$O(N \log N)$	$O(N)$	✓
CQRMF	a specially designed neural network	✓	★★☆☆☆	★★☆☆☆	$O(N \log N)$	$O(N)$	✓
CHR	Quantile regression model	✓	★★★★☆	★★★★☆	$O(N^2)$	$O(N^2)$	✓
LCP	Any model	✓	★★☆☆☆	★★☆☆☆	$O(N^2 \log N)$	$O(N^2)$	✓
CTI	Quantile regression model	✓	★★★★★	★★★★★	$O(N \log N)$	$O(MN)$	✗

Table 1: A comparison of different conformal regression methods in terms of resource consumption, performance, and other aspects.

the asymptotic time and space complexity as the calibration set size  $n$  approaches infinity, where  $M$  in the CTI framework represents the number of quantiles. In empirical evaluations, CTI, CHR, and LCP exhibited the longest computation times, followed by CQRMF and then CQRR (owing to its three-quantile structure). Notably, CQR-m demonstrated slightly increased latency compared to standard CQR due to the additional quantile computation, while Split maintained the fastest performance. The performance ranking generally followed this temporal hierarchy, with the exception of LCP: although this method serves as an enhancement to the Split approach, its localization mechanism can also be applied to improve other arbitrary models' performance.

## 4. Experiment

### 4.1. Simulation Study

Following the methodology outlined in [Luo and Zhou \(2025b\)](#), we generate the training data by drawing  $n = 10000$  independent, univariate predictor samples  $X_i$  from a uniform distribution on the interval  $[0, 1]$ <sup>1</sup>. The response variable  $Y_i$  is then sampled independently and identically distributed (i.i.d.) according to:

$$y \sim \text{Triangular}(0, x, x),$$

where  $\text{Triangular}(0, x, x)$  is the Triangular distribution with lower limit 0, upper limit  $x$ , and mode  $x$ . The conditional density is:

$$f(y|x) = \frac{2y}{x^2} \mathbb{1}\{y \in (0, x)\}. \quad (22)$$

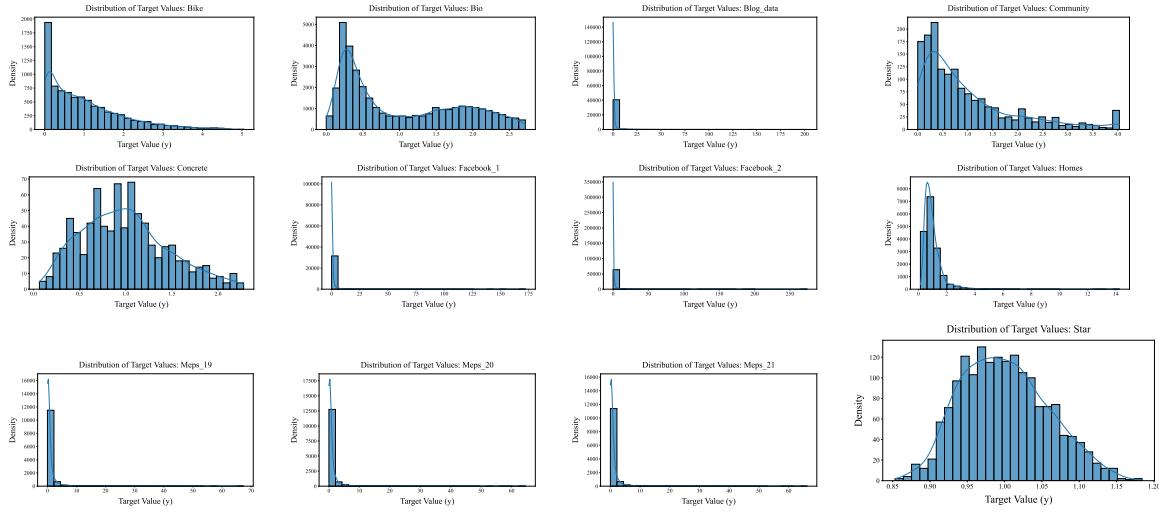
Figures 1-2 illustrates the distribution patterns of prediction intervals generated by various conformal regression methods in our simulation experiments. The color gradient from blue through orange, green, purple, to yellow indicates increasing frequency of interval overlaps. Distinct methodological differences can be visually discerned from these patterns.

First, the CTI method produces discrete prediction intervals, as evidenced by discontinuous matrix blocks in the visualization. Two key observations confirm this method's characteristics: 1) The prediction intervals are generally narrower, manifested through smaller overall matrix occupation in the plot; 2) Reduced presence of purple and orange regions indicates fewer areas with high overlap frequencies. Similar characteristics can be observed in the CHR method. Additionally, tail coverage analysis reveals higher performance in regions where  $Y < 0.1$  and  $Y > 0.9$ , as indicated by an increased number of red scatter points.

1. All the code for this experiment comes from the following links: <https://github.com/ml-stat-Sustech/TorchCP>, <https://github.com/luo-lorry/CTI>, and <https://github.com/LeyingGuan/LCP>. Additionally, all the visualization code for our experiments is sourced from: <https://github.com/bjbbbb/Conformal-Regression-Summarize>.

Name	Description	n	d
bike ( <a href="#">dat, a</a> )	bike sharing	10886	18
bio ( <a href="#">dat, b</a> )	physicochemical properties of protein tertiary structures	45730	9
blog ( <a href="#">dat, c</a> )	blog feedback	52397	280
community ( <a href="#">dat, d</a> )	community and crime	1994	100
concrete ( <a href="#">dat, e</a> )	concrete compressive strength	1030	8
facebook 1 ( <a href="#">dat, f</a> )	facebook comment volume	40948	53
facebook 2 ( <a href="#">dat, f</a> )	facebook comment volume	81311	53
homes ( <a href="#">dat, g</a> )	sale prices of homes in King County, Washington	21613	19
meps 19 ( <a href="#">dat, h</a> )	medical expenditure panel survey	15785	139
meps 20 ( <a href="#">dat, i</a> )	medical expenditure panel survey	17541	139
meps 21 ( <a href="#">dat, j</a> )	medical expenditure panel survey	15656	139
star ( <a href="#">Achilles et al., 2008</a> )	Tennessee's student-teacher achievement ratio	2161	39

Table 2: Dataset descriptions

Figure 3: The distribution plots of the target variable  $y$  for different datasets.

The CQR family exhibits distinct limitations: All variants generate unnecessarily wide intervals for  $x \in [0, 0.2]$ , resulting in increased interval lengths. Among these, CQR-r produces the largest intervals, while CQR-m and CQRFM demonstrate moderate improvements. Comparative analysis between CQR and CQR-m, as well as CQR-r and CQRFM, reveals that CQRM and CQRFM achieve significantly better tail coverage.

Finally, the LCP method - enhanced through split-based optimization of  $S_i^{split}$  - demonstrates the most efficient interval allocation, characterized by reduced green region density. This approach generates compact prediction intervals while maintaining superior tail coverage performance.

## 4.2. Real Data

Following the methodology outlined in [Sesia and Candès \(2020\)](#), we rescale the response  $Y$  by the mean absolute value. We randomly allocate 20% of the samples for testing, and from the remaining data, we utilize 70% for training the quantile regression model and 30% for calibration. This split has been validated in [Sesia and Candès \(2020\)](#). We repeat all experiments 50 times, starting from the initial data splitting. For the training procedure of quantile regression, except for

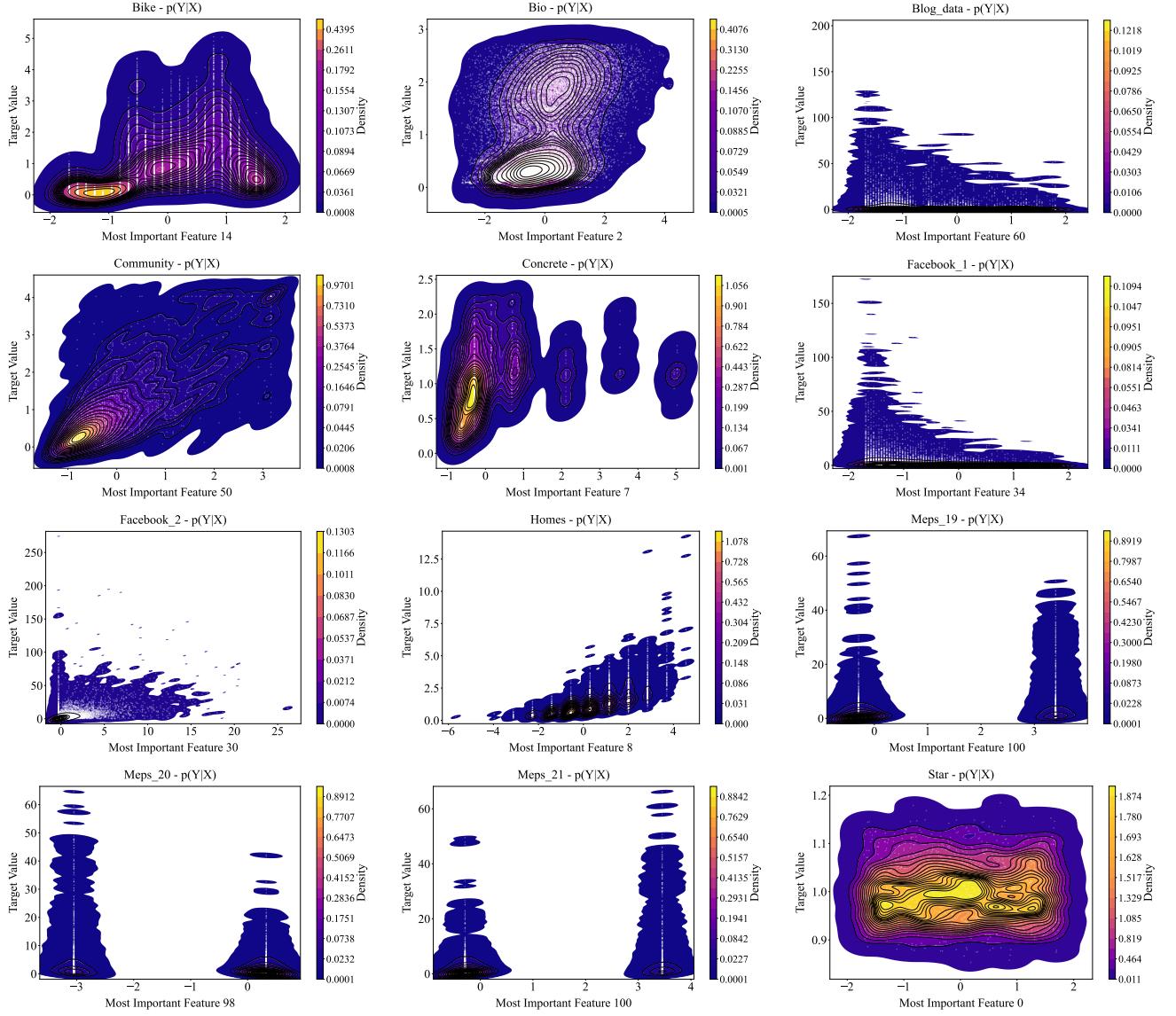


Figure 4: Given that the features vary across different datasets, to visualize the conditional density distributions, we employ a random forest to identify the most influential features for each dataset. Subsequently, we utilize two-dimensional Kernel Density Estimation (KDE) to plot the conditional distribution diagrams. In these diagrams, white scatter points denote the positions of actual values, with the x-axis representing the features and the y-axis indicating the target values. It can be observed that the distributions of most datasets are highly uneven. The meqs - series datasets exhibit similar distribution patterns, while the Facebook dataset and the blog dataset also share comparable distributions. In contrast, the distributions of the other datasets do not conform to a uniform pattern.

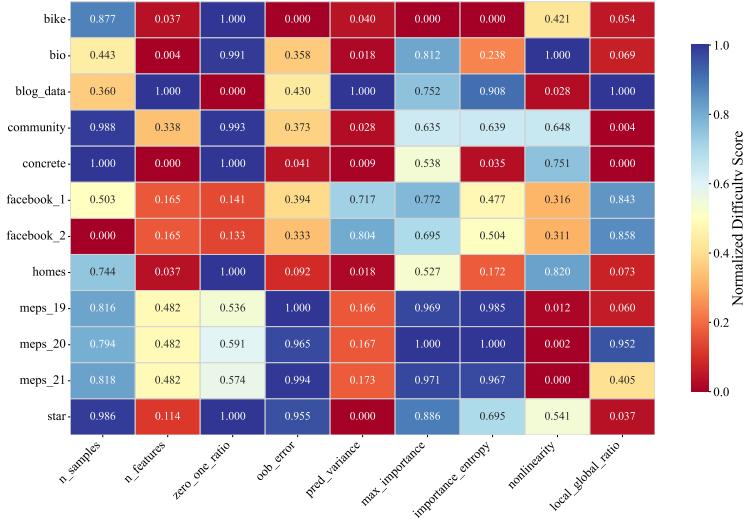


Figure 5: The heatmap visualizes the relative difficulty of regression datasets across nine normalized metrics. All metrics are scaled using min-max normalization, with three indicators (number of samples, maximum feature importance, and zero-one ratio in target) inverted to align higher values with greater difficulty. Metrics include: n\_samples , n\_features, zero\_one\_ratio, oob\_error, pred\_variance, max\_importance, importance\_entropy, nonlinearity, and local\_global\_ratio. Rows represent datasets, columns represent metrics, and color intensity reflects normalized difficulty scores (0: easiest, 1: hardest). Higher values in all metrics indicate increased dataset complexity for regression tasks.

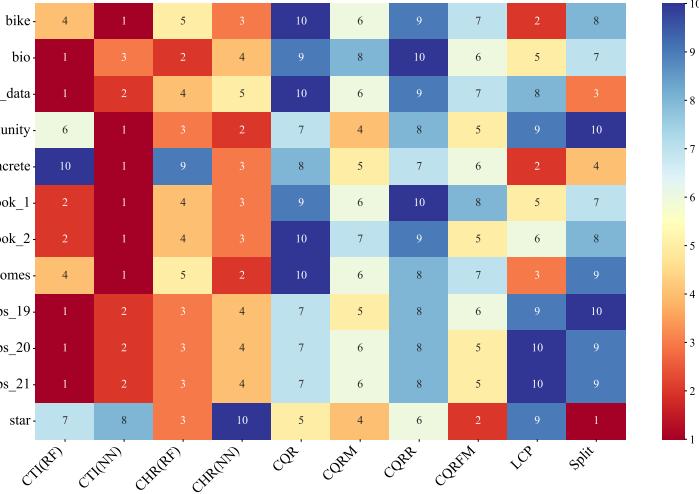


Figure 6: The heatmap presents the performance rankings of 10 regression methods across 12 real-world datasets based on prediction interval width (lower values indicate better performance). For each dataset, methods are ranked from 1 (best) to 10 (worst) based on ascending order of their interval width values, with tied ranks resolved by original precision. Rows represent datasets, columns represent methods, and numerical annotations show exact ranks.

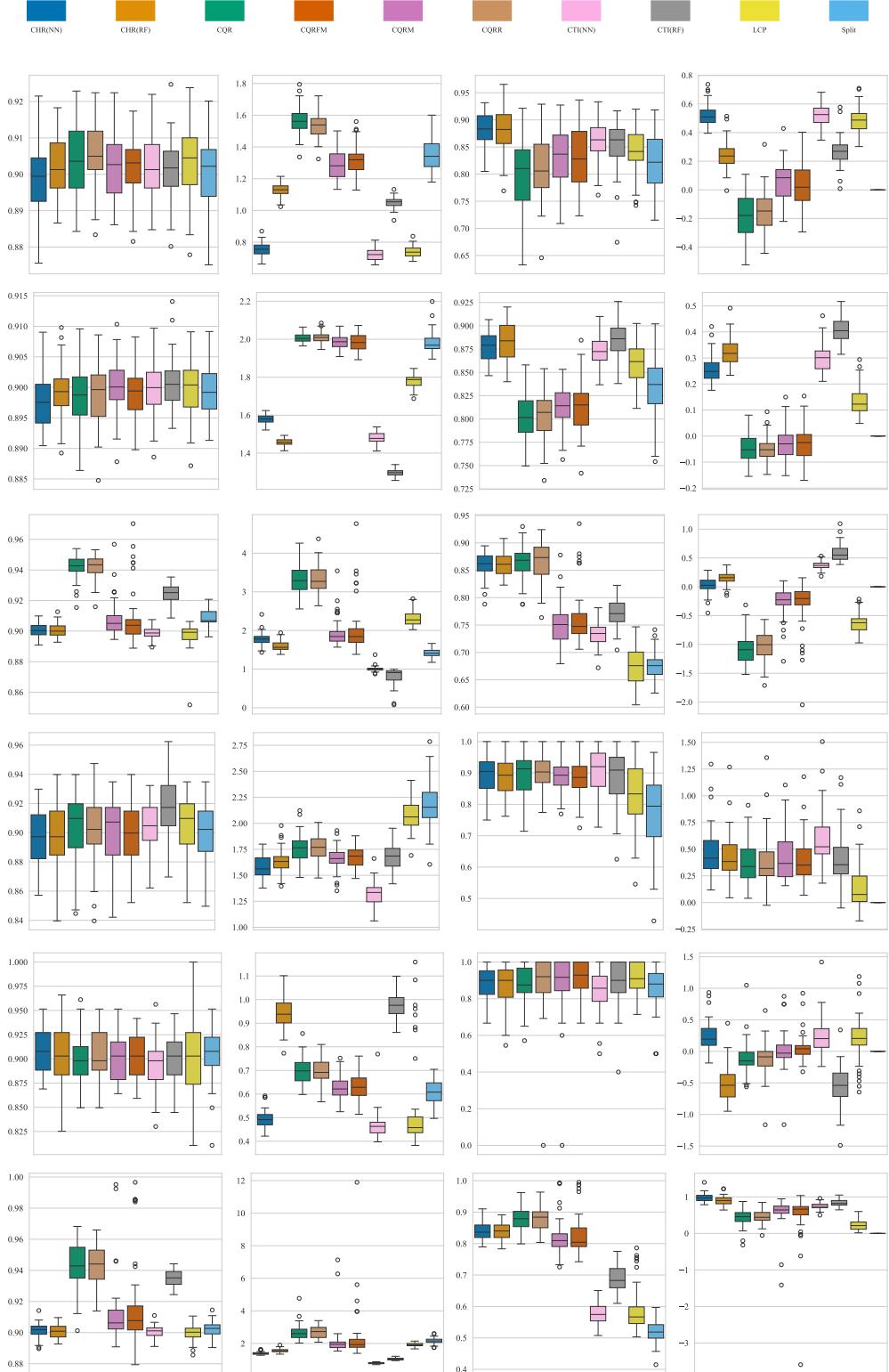


Figure 7: Box plots comparing various conformal regression methods across multiple datasets and evaluation metrics. Each experiment was conducted with 50 random dataset splits. The datasets are arranged vertically from top to bottom: bike, bio, blog, community, concrete, facebook 1, and facebook 2. The evaluation metrics, displayed horizontally from left to right, include Coverage, Size, BIS, and TCR.

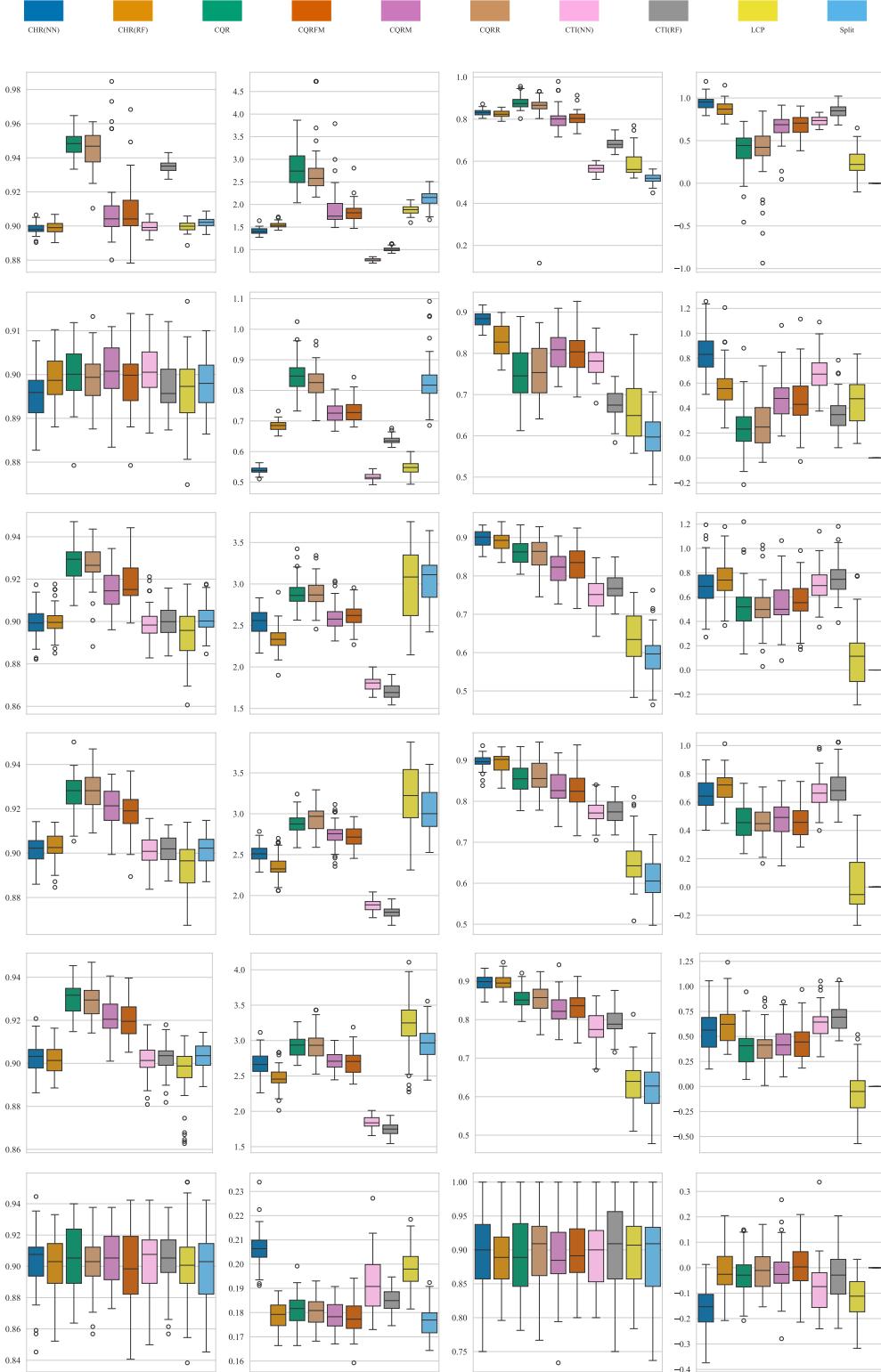


Figure 8: Box plots comparing various conformal regression methods across multiple datasets and evaluation metrics. Each experiment was conducted with 50 random dataset splits. The datasets are arranged vertically from top to bottom: facebook 2, homes, meps 19, meps 20, meps 21 and star. The evaluation metrics, displayed horizontally from left to right, include Coverage, Size, BIS, and TCR.

CTI and CHR (both of which employ both random forest (RF) and neural network (NN) models), all other methods utilize RF as the base predictive model. This setup allows us to both observe the impact of different base predictive models on the results and compare different conformal prediction methods.

Both CTI and CHR integrate the quantile regression results from Random Forests (RF) and Neural Networks (NN). For CQR, CQR-m, and CQR-r, we utilize the results from NN. For NFCP, we employ the Uniform strategy, and for LCP, we use the most basic standard residual score. We adopt four evaluation metrics, including coverage, size, and tail coverage rate (TCR) (Lin et al., 2021).

While individual metrics such as average interval Size and Worse-Slab Coverage (WSC) (Cauchois et al., 2021) are informative, they often reveal a fundamental trade-off: one method may yield narrower intervals at the cost of weaker conditional coverage, while another does the opposite. To facilitate a more holistic comparison that directly evaluates this balance, we introduce a heuristic metric for this study, the Balanced Interval Score (BIS). It is designed to provide a single, interpretable score for a method’s overall effectiveness. The BIS is formulated as:

$$\text{BIS} = \frac{\text{WSC}_{\text{method}}}{\text{WSC}_{\text{split}}} - \frac{\text{Size}_{\text{method}}}{\text{Size}_{\text{split}}}. \quad (23)$$

The rationale behind this formula is to normalize performance against a common baseline—the standard Split conformal method—and quantify the net gain. The term  $\text{WSC}_{\text{method}}/\text{WSC}_{\text{split}}$  measures the relative improvement in conditional coverage robustness, while  $\text{Size}_{\text{method}}/\text{Size}_{\text{split}}$  represents the relative cost in interval width. By subtracting the relative cost from the relative gain, the BIS captures the overall benefit a method provides over the simple baseline.

First, we will analyze 12 real-world datasets (as shown in Table 2). We have plotted the distribution of the true  $y$  values for these 12 datasets, as illustrated in Figure 3.

Additionally, as illustrated in Figure 4, we employed Kernel Density Estimation (KDE) to plot the conditional distribution graphs  $P(X | Y)$  for each dataset. Given that the number of features was not uniform across datasets, we utilized the random forest algorithm to select the most important feature for plotting purposes. Furthermore, we simultaneously plotted the normalized scores of several key indicators that significantly impact the predictive models across different datasets, as shown in Figure 5. It can be observed that the meps series datasets and the blog dataset are the most challenging to predict. Moreover, as depicted in Figure 6, we have also plotted the rankings of various CP methods based on their performance across different dataset sizes, thereby indicating which methods are more adept at handling specific types of datasets.

Table 3 presents the performance comparison of various conformal regression methods across one simulated dataset and twelve real-world datasets. Notably, the CTI method consistently maintained the smallest prediction interval size across most datasets, while TCR generally achieved the best overall performance. The second-ranked method, CHR, demonstrated superior adaptability on skewed datasets such as Facebook and MEPS. CQR exhibited stable WSC (Width of the Prediction Interval) values across all datasets, though its interval size tended to be larger on skewed data. The CQR method demonstrates favorable BIS performance on certain datasets but underperforms on others. Notably, its interval widths tend to be larger on skewed data. In contrast, CQR-m produces significantly narrower intervals than standard CQR across all datasets, while CQR-r exhibits intermediate performance between CQR and CQR-m with nearly identical BIS values across all three approaches. The CQRFM variant matches CQR-m’s interval dimensions while improving both TCR and BIS metrics. Compared to Split conformal regression, LCP shows consistent improvements in multiple evaluation metrics across most datasets, particularly for BIS. We further

Dataset	Metric	CTI(RF)	CTI(NN)	CHR(RF)	CHR(NN)	CQR	CQRM	CQRR	CQRFM	LCP	Split
synthetic	Coverage	0.900 (0.011)	0.898 (0.011)	0.897 (0.007)	0.900 (0.009)	0.900 (0.010)	0.901 (0.010)	0.899 (0.009)	0.899 (0.011)	0.916 (0.015)	0.900 (0.010)
	Size	<b>0.355 (0.007)</b>	0.363 (0.012)	0.361 (0.008)	0.369 (0.014)	0.442 (0.020)	0.402 (0.020)	0.434 (0.018)	0.403 (0.018)	0.430 (0.039)	0.475 (0.020)
	BIS	—	—	—	—	—	—	—	—	—	—
	TCR	<b>0.901 (0.015)</b>	0.843 (0.034)	0.850 (0.021)	0.833 (0.043)	0.722 (0.072)	0.804 (0.060)	0.744 (0.068)	0.776 (0.080)	0.798 (0.042)	0.764 (0.070)
bike	Coverage	0.901 (0.009)	0.902 (0.008)	0.902 (0.008)	0.900 (0.009)	0.904 (0.010)	0.903 (0.008)	0.905 (0.009)	0.901 (0.008)	0.903 (0.009)	0.900 (0.010)
	Size	1.049 (0.033)	<b>0.722 (0.036)</b>	1.128 (0.041)	0.757 (0.043)	1.564 (0.086)	1.289 (0.089)	1.534 (0.078)	1.313 (0.093)	0.740 (0.036)	1.355 (0.094)
	BIS	0.265 (0.010)	0.518 (0.006)	0.238 (0.001)	<b>0.520 (0.005)</b>	-0.187 (0.022)	0.060 (0.019)	-0.145 (0.022)	0.040 (0.023)	0.485 (0.009)	0.000 (0.000)
	TCR	<b>0.845 (0.022)</b>	0.828 (0.021)	0.837 (0.019)	0.844 (0.021)	0.621 (0.058)	0.7000 (0.055)	0.650 (0.060)	0.695 (0.048)	0.814 (0.029)	0.709 (0.041)
bio	Coverage	0.901 (0.004)	0.900 (0.005)	0.899 (0.004)	0.898 (0.005)	0.899 (0.005)	0.900 (0.004)	0.899 (0.005)	0.899 (0.004)	0.900 (0.005)	0.899 (0.004)
	Size	<b>1.297 (0.017)</b>	1.480 (0.028)	1.456 (0.019)	1.578 (0.022)	2.007 (0.024)	1.984 (0.034)	2.009 (0.029)	1.980 (0.041)	1.780 (0.032)	1.982 (0.053)
	BIS	<b>0.409 (0.002)</b>	0.302 (0.003)	0.257 (0.003)	0.326 (0.003)	-0.047 (0.025)	-0.025 (0.004)	-0.048 (0.004)	-0.023 (0.004)	0.135 (0.003)	0.000 (0.000)
	TCR	<b>0.862 (0.007)</b>	0.797 (0.010)	0.749 (0.010)	0.724 (0.012)	0.554 (0.026)	0.574 (0.028)	0.554 (0.025)	0.570 (0.027)	0.794 (0.016)	0.731 (0.033)
blog	Coverage	0.925 (0.006)	0.900 (0.003)	0.901 (0.004)	0.901 (0.005)	0.942 (0.008)	0.908 (0.011)	0.942 (0.008)	0.904 (0.018)	0.897 (0.008)	0.909 (0.005)
	Size	<b>0.797 (0.216)</b>	1.006 (0.070)	1.596 (0.125)	1.771 (0.171)	3.340 (0.359)	1.943 (0.360)	3.303 (0.360)	2.053 (0.648)	2.300 (0.197)	1.429 (0.105)
	BIS	<b>0.585 (0.022)</b>	0.377 (0.005)	0.154 (0.010)	0.028 (0.017)	-1.067 (0.070)	-0.265 (0.062)	-1.038 (0.035)	-0.294 (0.162)	-0.624 (0.030)	0.000 (0.000)
	TCR	<b>0.694 (0.221)</b>	0.631 (0.016)	0.266 (0.017)	0.249 (0.012)	0.251 (0.019)	0.573 (0.151)	0.297 (0.147)	0.652 (0.048)	0.617 (0.015)	0.101 (0.141)
community	Coverage	0.918 (0.020)	0.905 (0.016)	0.899 (0.022)	0.896 (0.019)	0.905 (0.023)	0.900 (0.021)	0.903 (0.022)	0.899 (0.023)	0.905 (0.019)	0.899 (0.021)
	Size	1.679 (0.119)	<b>1.325 (0.112)</b>	1.636 (0.116)	1.574 (0.104)	1.758 (0.133)	1.661 (0.116)	1.764 (0.116)	1.674 (0.104)	2.069 (0.142)	2.183 (0.224)
	BIS	0.410 (0.060)	<b>0.595 (0.060)</b>	0.430 (0.045)	0.464 (0.055)	0.368 (0.039)	0.418 (0.051)	0.386 (0.060)	0.402 (0.046)	0.152 (0.049)	0.000 (0.000)
	TCR	<b>0.839 (0.039)</b>	0.838 (0.044)	0.713 (0.053)	0.771 (0.041)	0.698 (0.094)	0.750 (0.064)	0.658 (0.102)	0.771 (0.066)	0.786 (0.050)	0.670 (0.062)
concrete	Coverage	0.901 (0.024)	0.895 (0.025)	0.903 (0.031)	0.907 (0.024)	0.900 (0.026)	0.902 (0.024)	0.904 (0.025)	0.903 (0.022)	0.898 (0.028)	0.905 (0.026)
	Size	0.974 (0.049)	<b>0.465 (0.055)</b>	0.942 (0.065)	0.493 (0.036)	0.700 (0.055)	0.628 (0.050)	0.699 (0.053)	0.629 (0.056)	0.466 (0.036)	0.607 (0.051)
	BIS	-0.555 (0.093)	0.243 (0.072)	-0.508 (0.065)	<b>0.243 (0.050)</b>	-0.108 (0.063)	0.014 (0.083)	-0.107 (0.062)	0.048 (0.047)	0.199 (0.126)	0.000 (0.000)
	TCR	0.770 (0.058)	<b>0.895 (0.057)</b>	0.725 (0.098)	0.880 (0.058)	0.743 (0.091)	0.799 (0.075)	0.768 (0.081)	0.811 (0.081)	0.866 (0.056)	0.802 (0.080)
facebook1	Coverage	0.935 (0.005)	0.901 (0.004)	0.901 (0.004)	0.901 (0.005)	0.943 (0.013)	0.913 (0.022)	0.944 (0.013)	0.918 (0.027)	0.900 (0.005)	0.902 (0.005)
	Size	1.048 (0.068)	<b>0.790 (0.045)</b>	1.559 (0.103)	1.389 (0.081)	2.699 (0.466)	2.148 (0.979)	2.696 (0.345)	2.367 (1.565)	1.925 (0.111)	2.156 (0.188)
	BIS	0.842 (0.009)	0.751 (0.008)	0.894 (0.015)	<b>0.975 (0.013)</b>	0.435 (0.050)	0.588 (0.148)	0.446 (0.032)	0.499 (0.420)	0.242 (0.028)	0.000 (0.000)
	TCR	0.755 (0.023)	0.667 (0.014)	0.356 (0.013)	0.339 (0.012)	0.418 (0.024)	0.835 (0.119)	0.458 (0.013)	<b>0.880 (0.038)</b>	0.671 (0.039)	0.540 (0.014)
facebook2	Coverage	0.935 (0.004)	0.900 (0.003)	0.899 (0.003)	0.899 (0.003)	0.949 (0.008)	0.910 (0.020)	0.935 (0.076)	0.900 (0.016)	0.900 (0.335)	0.902 (0.003)
	Size	1.014 (0.051)	<b>0.775 (0.033)</b>	1.547 (0.059)	1.413 (0.068)	2.778 (0.424)	1.910 (0.414)	2.747 (0.597)	1.847 (0.211)	1.883 (0.105)	2.140 (0.177)
	BIS	0.845 (0.005)	0.730 (0.003)	<b>0.943 (0.007)</b>	0.869 (0.007)	0.392 (0.046)	0.651 (0.024)	0.364 (0.015)	0.686 (0.014)	0.244 (0.022)	0.000 (0.000)
	TCR	0.750 (0.016)	0.662 (0.010)	0.345 (0.009)	0.344 (0.008)	0.421 (0.020)	0.825 (0.125)	0.456 (0.129)	<b>0.871 (0.021)</b>	0.665 (0.032)	0.532 (0.013)
homes	Coverage	0.898 (0.006)	0.900 (0.006)	0.900 (0.006)	0.895 (0.005)	0.900 (0.006)	0.901 (0.007)	0.899 (0.005)	0.899 (0.007)	0.896 (0.008)	0.897 (0.006)
	Size	0.638 (0.013)	<b>0.517 (0.010)</b>	0.684 (0.017)	0.538 (0.010)	0.847 (0.055)	0.728 (0.034)	0.826 (0.047)	0.734 (0.037)	0.545 (0.022)	0.829 (0.077)
	BIS	0.362 (0.021)	0.689 (0.022)	0.578 (0.034)	<b>0.836 (0.027)</b>	0.231 (0.035)	0.476 (0.031)	0.270 (0.036)	0.458 (0.041)	0.456 (0.034)	0.000 (0.000)
	TCR	0.715 (0.018)	0.785 (0.018)	0.797 (0.016)	<b>0.851 (0.014)</b>	0.680 (0.048)	0.726 (0.041)	0.696 (0.045)	0.718 (0.045)	0.733 (0.022)	0.640 (0.036)
meps19	Coverage	0.900 (0.007)	0.900 (0.008)	0.902 (0.007)	0.900 (0.007)	0.928 (0.009)	0.916 (0.010)	0.927 (0.010)	0.918 (0.009)	0.894 (0.012)	0.902 (0.008)
	Size	<b>1.703 (0.082)</b>	1.801 (0.084)	2.333 (0.163)	2.547 (0.151)	2.898 (0.177)	2.599 (0.168)	2.908 (0.190)	2.622 (0.146)	3.001 (0.389)	3.061 (0.274)
	BIS	0.756 (0.021)	0.696 (0.020)	0.749 (0.026)	0.696 (0.032)	0.521 (0.041)	0.541 (0.035)	0.512 (0.036)	0.557 (0.030)	0.107 (0.064)	0.000 (0.000)
	TCR	0.642 (0.020)	<b>0.655 (0.024)</b>	0.252 (0.020)	0.296 (0.018)	0.237 (0.028)	0.280 (0.147)	0.278 (0.130)	0.708 (0.022)	0.642 (0.027)	0.575 (0.031)
meps20	Coverage	0.902 (0.006)	0.901 (0.007)	0.902 (0.006)	0.902 (0.006)	0.927 (0.008)	0.920 (0.009)	0.928 (0.008)	0.918 (0.009)	0.894 (0.011)	0.902 (0.006)
	Size	<b>1.796 (0.070)</b>	1.882 (0.079)	2.357 (0.148)	2.515 (0.118)	2.879 (0.136)	2.745 (0.164)	2.942 (0.147)	2.718 (0.127)	3.231 (0.418)	3.052 (0.289)
	BIS	0.697 (0.017)	0.662 (0.015)	<b>0.705 (0.014)</b>	0.655 (0.013)	0.463 (0.016)	0.471 (0.014)	0.452 (0.017)	0.461 (0.012)	0.020 (0.040)	0.000 (0.000)
	TCR	0.638 (0.015)	<b>0.649 (0.018)</b>	0.237 (0.017)	0.282 (0.015)	0.227 (0.027)	0.277 (0.134)	0.238 (0.072)	0.696 (0.021)	0.637 (0.022)	0.573 (0.035)
meps21	Coverage	0.903 (0.007)	0.901 (0.007)	0.902 (0.007)	0.902 (0.008)	0.930 (0.007)	0.922 (0.009)	0.929 (0.007)	0.920 (0.008)	0.896 (0.013)	0.903 (0.007)
	Size	<b>1.751 (0.089)</b>	1.844 (0.083)	2.474 (0.156)	2.667 (0.179)	2.927 (0.149)	2.714 (0.134)	2.934 (0.196)	2.689 (0.164)	3.217 (0.403)	2.948 (0.245)
	BIS	<b>0.702 (0.024)</b>	0.640 (0.025)	0.627 (0.040)	0.559 (0.042)	0.403 (0.029)	0.424 (0.032)	0.397 (0.032)	0.448 (0.030)	-0.056 (0.064)	0.000 (0.000)
	TCR	0.638 (0.017)	<b>0.648 (0.019)</b>	0.234 (0.016)	0.289 (0.018)	0.211 (0.019)	0.239 (0.113)	0.219 (0.071)	0.685 (0.021)	0.642 (0.025)	0.567 (0.031)
star	Coverage	0.905 (0.018)	0.902 (0.020)	0.900 (0.019)	0.902 (0.019)	0.905 (0.020)	0.904 (0.018)	0.901 (0.017)	0.900 (0.023)	0.900 (0.022)	0.901 (0.022)
	Size	0.185 (0.005)	0.193 (0.011)	0.179 (0.006)	0.207 (0.007)	0.181 (0.006)	0.179 (0.006)	0.181 (0.006)	0.178 (0.007)	0.199 (0.008)	<b>0.177 (0.006)</b>
	BIS	-0.036 (0.008)	-0.083 (0.010)	-0.010 (0.008)	-0.159 (0.007)	-0.026 (0.007)	-0.013 (0.008)	-0.009 (0.006)	<b>0.001 (0.008)</b>	-0.114 (0.006)	0.000 (0.000)
	TCR	0.652 (0.075)	<b>0.713 (0.055)</b>	0.607 (0.071)	0.724 (0.059)	0.578 (0.079)	0.584 (0.077)	0.571 (0.070)	0.573 (0.090)	0.715 (0.076)	0.603 (0.090)

Table 3: The coverage, size, WSC and TCR results for various methods are presented in the table.

generated box plots (Figures 7-8) for four performance metrics across 12 real-world datasets, with each method tested under 50 random data partitions per dataset to ensure robustness.

**Choice of conformal regression method:** Based on the aforementioned extensive experiments, we hereby offer recommendations on how to select a CP method. First and foremost, a foundational predictive model is required, which holds particular significance for determining the size of the final prediction set. For datasets with a large number of x-features (more than 100), we recommend employing Random Forest as the foundational predictive model. Secondly, we suggest that if the generated prediction set does not require the characteristic of continuity, it is advisable to exclusively use CTI as the CP method. In cases where the dataset distribution is highly complex, we strongly recommend utilizing CHR as the CP method or combining methods such as CQR with LCP to enhance the model's performance. Finally, when dealing with datasets that have a very small sample size or a simple distribution, we recommend using the Split method.

## 5. Conclusion

This study provides a comprehensive review and in-depth comparative analysis of univariate conformal regression methods, systematically investigating their principles, performance, and re-

source consumption. Our empirical results demonstrate pronounced discrepancies across different techniques in critical metrics such as prediction interval width, coverage probability, and tail coverage performance. While this work establishes a significant benchmark, we acknowledge that its scope is intentionally centered on the popular split conformal prediction framework, valued for its computational efficiency and widespread adoption. This focus naturally illuminates several promising avenues for future research. A crucial next step is to extend this comparative analysis to other foundational paradigms, including full conformal, cross-conformal, and Jackknife+ procedures, which offer different trade-offs between statistical efficiency and computational cost. Beyond this, further progress can be made by pursuing the deep integration of conformal prediction with model training pipelines, developing efficient inference frameworks for resource-constrained environments, extending these paradigms to complex real-world scenarios, and realizing improved conditional coverage guarantees through novel methodological formulations.

## Acknowledgments

The work described in this paper was partially supported by grants from City University of Hong Kong (9610639, 6000864) and Chengdu Municipal Office of Philosophy and Social Science (2024BS013).

## References

- Bike sharing dataset data set. <https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset>, a. Accessed: July, 2019.
- Physicochemical properties of protein tertiary structure data set. <https://archive.ics.uci.edu/ml/datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure>, b. Accessed: July, 2019.
- BlogFeedback data set. <https://archive.ics.uci.edu/ml/datasets/BlogFeedback>, c. Accessed: July, 2019.
- Communities and crime data set. <http://archive.ics.uci.edu/ml/datasets/communities+and+crime>, d. Accessed: July, 2019.
- Concrete compressive strength data set. <http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength>, e. Accessed: July, 2019.
- Facebook comment volume data set. <https://archive.ics.uci.edu/ml/datasets/Facebook+Comment+Volume+Dataset>, f. Accessed: July, 2019.
- House sales in King County, USA. <https://www.kaggle.com/harlfoxem/housesalesprediction/metadata>, g. Accessed: August, 2019.
- Medical expenditure panel survey, panel 19. [https://meps.ahrq.gov/mepsweb/data\\_stats/download\\_data\\_files\\_detail.jsp?cboPufNumber=HC-181](https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-181), h. Accessed: July, 2019.
- Medical expenditure panel survey, panel 20. [https://meps.ahrq.gov/mepsweb/data\\_stats/download\\_data\\_files\\_detail.jsp?cboPufNumber=HC-181](https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-181), i. Accessed: July, 2019.
- Medical expenditure panel survey, panel 21. [https://meps.ahrq.gov/mepsweb/data\\_stats/download\\_data\\_files\\_detail.jsp?cboPufNumber=HC-192](https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-192), j. Accessed: July, 2019.

C.M. Achilles, Helen Pate Bain, Fred Bellott, Jayne Boyd-Zaharias, Jeremy Finn, John Folger, John Johnston, and Elizabeth Word. Tennessee's student teacher achievement ratio (STAR) project, 2008. URL <https://doi.org/10.7910/DVN/SIWH9F>. Accessed: July, 2019.

Anastasios N Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. *arXiv preprint arXiv:2208.02814*, 2022.

Anastasios N Angelopoulos, Stephen Bates, et al. Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16(4):494–591, 2023.

Jie Bao, Chuangyin Dang, Rui Luo, Hanwei Zhang, and Zhixin Zhou. Enhancing adversarial robustness with conformal prediction: A framework for guaranteed model reliability. In *Proceedings of the Forty-second International Conference on Machine Learning (ICML)*, 2025.

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1):486–507, 2021.

Anthony Bellotti. Optimized conformal classification using gradient descent approximation. *arXiv preprint arXiv:2105.11255*, 2021.

Henrik Boström, Henrik Linusson, Tuve Löfström, and Ulf Johansson. Accelerating difficulty estimation for conformal regression forests. *Annals of Mathematics and Artificial Intelligence*, 81: 125–144, 2017.

Maxime Cauchois, Suyash Gupta, and John C Duchi. Knowing what you know: valid and validated confidence sets in multiclass and multilabel prediction. *Journal of machine learning research*, 22(81):1–42, 2021.

Wenyu Chen, Kelli-Jean Chun, and Rina Foygel Barber. Discretized conformal prediction for efficient distribution-free inference. *Stat*, 7(1):e173, 2018.

Xi Chen, Rahul Bhadani, and Larry Head. Conformal trajectory prediction with multi-view data integration in cooperative driving. *arXiv preprint arXiv:2408.00374*, 2024.

Matt Y Cheung, Tucker J Netherton, Laurence E Court, Ashok Veeraraghavan, and Guha Balakrishnan. Regression conformal prediction under bias. *arXiv preprint arXiv:2410.05263*, 2024.

Hugh A Chipman, Edward I George, and Robert E McCulloch. Bart: Bayesian additive regression trees. *Annals of Applied Statistics*, 4(1):266–298, 2010.

Nicolo Colombo. On training locally adaptive cp. In *Conformal and Probabilistic Prediction with Applications*, pages 384–398. PMLR, 2023.

Nicolo Colombo. Normalizing flows for conformal regression. *arXiv preprint arXiv:2406.03346*, 2024.

Nathaniel Diamant, Ehsan Hajiramezanali, Tommaso Biancalani, and Gabriele Scalia. Conformalized deep splines for optimal and efficient prediction sets. In *International Conference on Artificial Intelligence and Statistics*, pages 1657–1665. PMLR, 2024.

Shian Du, Yihong Luo, Wei Chen, Jian Xu, and Delu Zeng. To-flow: Efficient continuous normalizing flows with temporal optimization adjoint with moving speed. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12570–12580, 2022.

Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

Ruijiang Gao, Mingzhang Yin, James Mcinerney, and Nathan Kallus. Adjusting regression models for conditional uncertainty calibration. *Machine Learning*, pages 1–24, 2024.

Isaac Gibbs, John J Cherian, and Emmanuel J Candès. Conformal prediction with conditional guarantees. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkaf008, 2025.

Natalia Martinez Gil, Dhaval Patel, Chandra Reddy, Giridhar Ganapavarapu, Roman Vaculin, and Jayant Kalagnanam. Identifying homogeneous and interpretable groups for conformal prediction. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, pages 2471–2485, 2024.

Leying Guan. Localized conformal prediction: A generalized inference framework for conformal prediction. *Biometrika*, 110(1):33–50, 2023.

Rohan Hore and Rina Foygel Barber. Conformal prediction with local weights: randomization enables robust guarantees. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(2):549–578, 2025.

Jianguo Huang, Jianqing Song, Xuanning Zhou, Bingyi Jing, and Hongxin Wei. Torchcp: A python library for conformal prediction, 2024.

Pierre Humbert, Batiste Le Bars, Aurélien Bellet, and Sylvain Arlot. One-shot federated conformal prediction. In *International Conference on Machine Learning*, pages 14153–14177. PMLR, 2023.

Rafael Izbicki, Gilson T Shimizu, and Rafael B Stern. Flexible distribution-free conditional predictive bands using density estimators. *arXiv preprint arXiv:1910.05575*, 2019.

Ulf Johansson, Henrik Boström, Tuve Löfström, and Henrik Linusson. Regression conformal prediction with random forests. *Machine learning*, 97:155–176, 2014a.

Ulf Johansson, Cecilia Sönströd, Henrik Linusson, and Henrik Boström. Regression trees for streaming data with local performance guarantees. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 461–470. IEEE, 2014b.

Ulf Johansson, Cecilia Sönströd, and Henrik Linusson. Efficient conformal regressors using bagged neural nets. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2015.

Danijel Kvaranovic, Kory D Johnson, and Hannes Leeb. Adaptive, distribution-free prediction intervals for deep networks. In *International Conference on Artificial Intelligence and Statistics*, pages 4346–4356. PMLR, 2020.

Shayan Kiyani, George Pappas, and Hamed Hassani. Conformal prediction with learned features. *arXiv preprint arXiv:2404.17487*, 2024a.

Shayan Kiyani, George J Pappas, and Hamed Hassani. Length optimization in conformal prediction. *Advances in Neural Information Processing Systems*, 37:99519–99563, 2024b.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):71–96, 2014.

- Jing Lei, James Robins, and Larry Wasserman. Distribution-free prediction sets. *Journal of the American Statistical Association*, 108(501):278–287, 2013.
- Jing Lei, Max G’Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.
- Zhen Lin, Shubhendu Trivedi, and Jimeng Sun. Locally valid and discriminative prediction intervals for deep learning models. *Advances in Neural Information Processing Systems*, 34:8378–8391, 2021.
- Lars Lindemann, Matthew Cleaveland, Gihyun Shim, and George J Pappas. Safe planning in dynamic environments using conformal prediction. *IEEE Robotics and Automation Letters*, 8(8):5116–5123, 2023.
- Rui Luo and Nicolo Colombo. Entropy reweighted conformal classification. In *The 13th Symposium on Conformal and Probabilistic Prediction with Applications*, pages 264–276. PMLR, 2024.
- Rui Luo and Zhixin Zhou. Trustworthy classification through rank-based conformal prediction sets. *arXiv preprint arXiv:2407.04407*, 2024.
- Rui Luo and Zhixin Zhou. Conditional conformal risk adaptation. *arXiv preprint arXiv:2504.07611*, 2025a.
- Rui Luo and Zhixin Zhou. Conformal thresholded intervals for efficient regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(18):19216–19223, 2025b.
- Rui Luo and Zhixin Zhou. Conformalized interval arithmetic with symmetric calibration. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(18):19207–19215, 2025c.
- Rui Luo and Zhixin Zhou. Conformity score averaging for classification. In *Forty-second International Conference on Machine Learning*, 2025d.
- Rui Luo and Zhixin Zhou. Conformal thresholded intervals for efficient regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(18):19216–19223, 2025e.
- Rui Luo and Zhixin Zhou. Volume-sorted prediction set: Efficient conformal prediction for multi-target regression. *arXiv preprint arXiv:2503.02205*, 2025f.
- Rui Luo, Buddhika Nettasinghe, and Vikram Krishnamurthy. Anomalous edge detection in edge exchangeable social network models. In *Conformal and probabilistic prediction with applications*, pages 287–310. PMLR, 2023.
- Rui Luo, Jie Bao, Zhixin Zhou, and Chuangyin Dang. Game-theoretic defenses for robust conformal prediction against adversarial attacks in medical imaging. *arXiv preprint arXiv:2411.04376*, 2024.
- Malik Magdon-Ismail and Amir Atiya. Neural networks for density estimation. *Advances in Neural Information Processing Systems*, 11, 1998.
- Nicolai Meinshausen and Greg Ridgeway. Quantile regression forests. *Journal of machine learning research*, 7(6), 2006.
- Julia A Meister and Khuong An Nguyen. Conformalised data synthesis. *Machine Learning*, 114(3):1–37, 2025.

Sang Jun Moon, Jong-June Jeon, Jason Sang Hun Lee, and Yongdai Kim. Learning multiple quantiles with neural networks. *Journal of Computational and Graphical Statistics*, 30(4):1238–1248, 2021.

William Overman, Jacqueline Vallon, and Mohsen Bayati. Aligning model properties via conformal risk control. *Advances in Neural Information Processing Systems*, 37:110702–110722, 2024.

Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *Machine learning: ECML 2002: 13th European conference on machine learning Helsinki, Finland, August 19–23, 2002 proceedings 13*, pages 345–356. Springer, 2002.

Harris Papadopoulos, Alex Gammerman, and Volodya Vovk. Normalized nonconformity measures for regression conformal prediction. In *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA 2008)*, pages 64–69, 2008.

Harris Papadopoulos, Vladimir Vovk, and Alex Gammerman. Regression conformal prediction with nearest neighbours. *Journal of Artificial Intelligence Research*, 40:815–840, 2011.

Vincent Plassier, Alexander Fishkov, Mohsen Guizani, Maxim Panov, and Eric Moulines. Probabilistic conformal prediction with approximate conditional validity. *arXiv preprint arXiv:2407.01794*, 2024.

Thomas Pouplin, Alan Jeffares, Nabeel Seedat, and Mihaela Van Der Schaar. Relaxed quantile regression: prediction intervals for asymmetric noise. *arXiv preprint arXiv:2406.03258*, 2024.

Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in neural information processing systems*, 32, 2019.

Aviv A Rosenberg, Sanketh Vedula, Yaniv Romano, and Alex M Bronstein. Fast nonlinear vector quantile regression. *arXiv preprint arXiv:2205.14977*, 2022.

Matteo Sesia and Emmanuel J Candès. A comparison of some conformal quantile regression methods. *Stat*, 9(1):e261, 2020.

Matteo Sesia and Yaniv Romano. Conformal prediction using conditional histograms. *Advances in Neural Information Processing Systems*, 34:6304–6315, 2021.

Matteo Sesia, Stefano Favaro, and Edgar Dobriban. Conformal frequency estimation using discrete sketched data with coverage for distinct queries. *Journal of Machine Learning Research*, 24(348):1–80, 2023.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

Martim Sousa, Ana Maria Tomé, and José Moreira. Improving conformalized quantile regression through cluster-based feature relevance. *Expert Systems with Applications*, 238:122322, 2024.

David Stutz, Ali Taylan Cemgil, Arnaud Doucet, et al. Learning optimal conformal classifiers. *International Conference on Learning Representations (ICLR)*, 2022.

Xiaoyi Su, Zhixin Zhou, and Rui Luo. Adaptive conformal inference by particle filtering under hidden markov models. *arXiv preprint arXiv:2411.01558*, 2024.

- Lingxuan Tang, Rui Luo, Zhixin Zhou, and Nicolo Colombo. Enhanced route planning with calibrated uncertainty set. *Machine Learning*, 114(5):1–16, 2025.
- Lars van der Laan and Ahmed M Alaa. Self-calibrating conformal prediction. *arXiv preprint arXiv:2402.07307*, 2024.
- Janette Vazquez and Julio C Facelli. Conformal prediction in clinical medical sciences. *Journal of Healthcare Informatics Research*, 6(3):241–252, 2022.
- Vladimir Vovk. Cross-conformal predictors. *Annals of Mathematics and Artificial Intelligence*, 74:9–28, 2015.
- Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.
- Vladimir Vovk, Ilia Nouretdinov, and Alex Gammerman. On-line predictive linear regression. *The Annals of Statistics*, pages 1566–1590, 2009.
- Vladimir Vovk, Ivan Petej, Paolo Tocacceli, Alexander Gammerman, Ernst Ahlberg, and Lars Carlsson. Conformal calibrators. In *conformal and probabilistic prediction and applications*, pages 84–99. PMLR, 2020.
- Volodya Vovk, Alexander Gammerman, and Craig Saunders. Machine-learning applications of algorithmic randomness. In *International Conference on Machine Learning*, 1999.
- Ting Wang, Zhixin Zhou, and Rui Luo. Enhancing trustworthiness of graph neural networks with rank-based conformal training. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(20):21261–21268, 2025.
- Jerzy Wieczorek. Design-based conformal prediction. *arXiv preprint arXiv:2303.01422*, 2023.
- Margaux Zaffran, Aymeric Dieuleveut, Julie Josse, and Yaniv Romano. Conformal prediction with missing values. In *International Conference on Machine Learning*, pages 40578–40604. PMLR, 2023.
- Zheng Zhang, Jie Bao, Zhixin Zhou, Nicolo Colombo, Lixin Cheng, and Rui Luo. Residual reweighted conformal prediction for graph neural networks. In *Proceedings of the 41st Conference on Uncertainty in Artificial Intelligence (UAI)*, 2025. to appear.