SPACR: Single-Pass Adaptive Training of Uncertainty-Aware Conformal Regressors

Anonymous authors
Paper under double-blind review

Abstract

Conformal Prediction (CP) provides robust uncertainty guarantees for predictive models, but is typically applied post hoc, which misaligns model training with the conformal goal of producing efficient (i.e, narrow) intervals. We propose SPACR (Single-Pass Adaptive Conformal Regressor), a novel method for directly training uncertainty-aware regressors within a differentiable loss. SPACR jointly optimizes efficiency and validity without batch-splitting or a predefined confidence levels during training. As a result, a single SPACR model yields valid prediction intervals at multiple confidence levels during inference, avoiding the costly retraining required by methods like DOICR. Experiments on diverse datasets show that SPACR consistently gives tighter intervals and better coverage-efficiency trade-offs compared to standard CP and DOICR, while significantly reducing computational costs.

1 Introduction

Quantifying uncertainty is a key challenge in Machine Learning (ML), particularly where incorrect predictions can lead to serious consequences. In healthcare, a wrong diagnosis can harm patients (Vazquez & Facelli (2022); Seoni et al. (2023)), while in autonomous driving, mistakes can put pedestrians and drivers at risk (Michelmore et al. (2020)). In environmental monitoring, uncertainty-aware methods are vital in natural disasters for faster responses to protect people and property (Han & Coulibaly (2017); Allaire et al. (2021)). These high-stakes fields increasingly require models to not only produce accurate predictions but also quantify their reliability. This enables experts to step in when needed, thereby creating a collaborative human-AI decision-making framework instead of a blind reliance on overconfident but incorrect predictions.

Conformal Prediction (CP) is a powerful distribution-free framework for generating uncertainty estimates, popular for its theoretical guarantees and model-agnostic nature (Vovk et al. (2005); Angelopoulos & Bates (2021)). Instead of single point predictions, CP produces sets (for classification) or intervals (for regression) that come with statistical guarantees. By selecting a confidence level, users can control the probability that the true value falls within the predicted range, offering a robust and interpretable measure of uncertainty. Unlike baselines such as Deep Ensembles (Lakshminarayanan et al. (2017)) or Bayesian deep learning that explicitly model aleatoric and epistemic uncertainty (Kendall & Gal (2017)), CP requires no assumptions beyond data exchangeability, making it easily and robustly applicable to diverse ML models.

While CP has demonstrated significant potential for uncertainty quantification, a common limitation of conventional conformal methods such as Inductive Conformal Prediction (ICP) remains: these frameworks are typically applied *post hoc*, treating the underlying model as a "black box". Thus, the optimization objective of this model is often misaligned with the conformal goal, which is to maximize predictive efficiency (i.e., the smallest possible confidence sets or intervals) that still guarantee valid coverage. This means that conventional CP approaches can lead to unnecessarily wide intervals and less effective uncertainty estimates.

This challenge has motivated the development of methods that directly train models to maximize predictive efficiency. For classification, Colombo & Vovk (2020) proposed a direct approach to training CP, where the model learns a probabilistic measure of efficiency through optimization. However, this approach involves multiple step functions, which are non-differentiable. To overcome this, Bellotti (2021) suggested approximating the CP goal with a differentiable function that can be optimized using gradient descent. Stutz et al.

(2022) further improved this work by accounting for the calibration step during training, which plays a key role in reducing inefficiency. The broader shift toward differentiable conformal objectives is further evidenced by recent frameworks exploring the end-to-end optimization of conformal risk control (Yeh et al. (2025)).

Although these studies are interesting, they do not tackle regression. To our knowledge, Lei & Bellotti (2023) is the only work that extends conformal training to regression. Their method, Directly Optimized Inductive Conformal Regression (DOICR), minimizes ICP interval width while enforcing validity via a Deep Neural Network loss. However, DOICR requires explicit batch splitting and a predefined confidence level during training. As a result, it must batch-split data to incorporate calibration, and should be retrained for each different confidence level, which limits its sample efficiency and increases computational costs.

To address these limitations, we propose a new framework: *Single-Pass Adaptive Conformal Regressor (SPACR)*. Its novelty lies in integrating known components (accuracy, efficiency, and validity) into a unified, differentiable training objective. This simultaneously ensures calibration and efficiency without requiring batch-splitting or a fixed confidence level during training. This results in a single model that can produce intervals at multiple confidence levels during inference. To our knowledge, SPACR is the first method to enable such single-pass adaptive conformal regression with these capabilities. Our contributions are:

- We introduce SPACR, a differentiable framework for training uncertainty-aware regressors by directly embedding conformal objectives into the loss function.
- We propose a unified optimization strategy that simultaneously minimizes point prediction error, penalizes interval width for efficiency, and enforces a validity penalty to ensure reliable coverage.
- SPACR mathematically decouples its uncertainty estimate from the target confidence level during training. This eliminates the sample-inefficient batch-splitting required by DOICR, enabling a single model to adaptively generate valid intervals across any confidence level during inference.
- We empirically demonstrate that SPACR achieves consistently tighter intervals and superior coverage-efficiency trade-offs compared to standard CP methods and state-of-the-art baselines like DOICR, while significantly reducing computational costs.

The remainder of this paper is organized as follows: Section 2 provides the background on Conformal Prediction and related work in training uncertainty-aware regressors. Section 3 details the SPACR architecture and its unified optimization objective. Sections 4 and 5 outline the experimental setup and evaluate our method against existing baselines. Finally, Section 6 concludes and discusses future work.

2 Related work

2.1 Conformal Prediction for Regression

Conformal Prediction (CP) (Vovk et al. (2005); Angelopoulos & Bates (2021)) is a model-agnostic framework that provides distribution-free uncertainty guarantees. For regression, CP constructs prediction intervals $\hat{C}_{1-\alpha}(x)$ that satisfy marginal coverage : $\mathbb{P}\big(y \in \hat{C}_{1-\alpha}(x)\big) \geq 1-\alpha$, where $\alpha \in (0,1)$ is the target error rate. To avoid the impractical $\mathcal{O}(n)$ per-prediction cost of Transductive CP, Papadopoulos (2008) introduced Inductive CP (ICP) as an efficient alternative that adopts a split-conformal approach to reduce prediction complexity to $\mathcal{O}(1)$, making ICP the practical standard.

ICP uses an exchangeable dataset partitioned into a proper training set $\mathcal{D}_{train} = \{(x_i, y_i)\}_{i=1}^{n}$ and a calibration set $\mathcal{D}_{cal} = \{(x_j, y_j)\}_{j=1}^{n_{cal}}$, where each input $x \in \mathcal{X}$ is associated with a real-valued target $y \in \mathbb{R}$. A regression model $f : \mathcal{X} \to \mathbb{R}$ is trained on $\mathcal{D}_{train}$. Next, standard non-conformity scores s_j are computed on $\mathcal{D}_{cal}$ to measure how much an example deviates from the model's predictions, with:

$$s_j = |y_j - f(x_j)|. \tag{1}$$

Using the $(1-\alpha)$-quantile of these scores as the threshold $\hat{q}$, the prediction interval for a new point x_{test} is:

$$\hat{C}_{1-\alpha}(x_{test}) = \Big[f(x_{test}) - \hat{q},\ f(x_{test}) + \hat{q}\Big]. \tag{2}$$

For a normalized method, Papadopoulos & Haralambous (2011) calculate scores on $\mathcal{D}_{cal}$ as:

$$s_j = \frac{|y_j - f(x_j)|}{\sigma_j}, \tag{3}$$

where σ_j estimates the instance-specific difficulty. For a new input x_{test}, the final prediction interval becomes:

$$\hat{C}_{1-\alpha}(x_{test}) = \Big[f(x_{test}) - \hat{q}\sigma_{test},\ f(x_{test}) + \hat{q}\sigma_{test}\Big]. \tag{4}$$

The normalization term σ enables adaptive intervals: simple instances with small σ get tighter bounds, while challenging cases receive wider intervals. This generalizes early ICP implementations that used fixed-width intervals ($\sigma \equiv 1$), demonstrating how modern approaches improve efficiency without sacrificing coverage.

While ICP guarantees marginal coverage under exchangeability and uses σ for efficiency, its required data-splitting reduces effective training data. To address this, particularly for small datasets, end-to-end conformal learning integrates calibration into training to simultaneously optimize accuracy and interval width.

2.2 Training Conformal Regressors

To overcome ICP limitations, recent methods directly train models to produce well-calibrated uncertainty estimates. This aligns standard training objectives (like point prediction accuracy) with the conformal goals of maximizing predictive efficiency (i.e., minimizing the size of prediction intervals) while ensuring validity.

Lei & Bellotti (2023) presented Directly Optimized Inductive Conformal Regression (DOICR), which extends conformal training (Stutz et al. (2022)) from classification to regression. DOICR aims to directly optimize ICP by minimizing the prediction intervals' average width while maintaining the desired coverage. This is achieved by incorporating a specialized loss function into the training of a Deep Neural Network (DNN).

DOICR relies on an iterative training procedure where, in each epoch, $\mathcal{D}_{train}$ is randomly split into a primary training set ($\mathcal{D}_1$) and an embedded calibration set ($\mathcal{D}_2$). The DNN outputs both a predicted mean $m(x)$ and a log-uncertainty $u(x)$. For $(x_j, y_j) \in \mathcal{D}_2$, normalized non-conformity scores are computed as $s_j = \frac{|y_j - m(x_j)|}{\exp(u(x_j))}$. The $(1-\alpha)$-quantile q of these defines the conformal threshold, which is integrated into the loss function as:

$$\mathcal{L}_{\text{DOICR}} = \frac{2q}{|\mathcal{D}_1|} \sum_{(x_i, y_i) \in \mathcal{D}_1} \exp(u(x_i)),$$

which combines the quantile q with the sum of $\exp(u(x_i))$ over the primary training set $\mathcal{D}_1$. By back-propagating through this loss, DOICR jointly optimizes predictions and uncertainty estimates. However, the reliance on batch-wise data splitting, which seems less sample efficient, can lead to issues when there is limited access to training data. Also, it requires a fixed significance level α during training through the quantile q. Consequently, various α values demand retraining, which increases the computational complexity.

3 SPACR : Single-Pass Adaptive Conformal Regressor

Addressing the limitations of prior approaches, we introduce SPACR, a novel framework that integrates ICP into the training loop without data splitting and without fixing a significance level during training.

Given a proper training set $\mathcal{D}_{train} = \{(x_i, y_i)\}_{i=1}^n$, our regressor $f_\theta : \mathcal{X} \to \mathbb{R} \times \mathbb{R}$ outputs both a predicted mean $\hat{y}_i$ and a raw uncertainty score u_i for each input x_i. The prediction interval during training is then:

$$[\hat{y}_i - \hat{\sigma}_i, \hat{y}_i + \hat{\sigma}_i],$$

where the uncertainty term $\hat{\sigma}_i$ dictates the interval size and is parameterized through an exponential mapping:

$$\hat{\sigma}_i = \exp(u_i) \quad \text{with} \quad u_i \in \mathbb{R}.$$

This mapping ensures positive interval widths, stable gradients, and adaptation to input difficulty (narrow intervals for predictable samples and wide intervals for uncertain ones) implemented end-to-end.

To train SPACR, we propose a custom loss function that jointly optimizes three objectives:

- **Accuracy ($\mathcal{L}_{\textbf{Accuracy}}$)**: Measured by Mean Absolute Error (MAE) for precise point predictions.
- **Efficiency ($\mathcal{L}_{\textbf{Efficiency}}$)**: A term that penalizes large uncertainties to encourage narrow intervals.
- **Validity ($\mathcal{L}_{\textbf{Validity}}$)**: A loss that imposes a penalty when the true value falls outside the predicted interval, thus forcing the model to produce genuine interval predictions rather than point predictions.

The overall loss is defined as:

$$\mathcal{L}_{\text{SPACR}} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|}_{\mathcal{L}_{\text{Accuracy}}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_i}_{\mathcal{L}_{\text{Efficiency}}} + \underbrace{\lambda\left[\frac{1}{n}\sum_{i=1}^{n}\phi(y_i, \hat{y}_i, \hat{\sigma}_i)\right]}_{\mathcal{L}_{\text{Validity}}}, \tag{5}$$

where $\lambda > 0$ controls the importance of the validity term in the overall loss, and $\phi(\cdot)$ computes the distance the true target falls outside the predicted bounds, given by $\phi(y, \hat{y}, \hat{\sigma}) = \max(|y - \hat{y}| - \hat{\sigma}, 0)$.

Since $\mathcal{L}_{\text{Accuracy}}$ and $\mathcal{L}_{\text{Efficiency}}$ share the same scale (units of y), they are weighted equally. This leaves only λ to control coverage, restricting the hyperparameter search space and preserving computational efficiency.

As illustrated in Figure 1, this loss encourages the model to dynamically adjust interval widths based on empirical coverage violations, which aligns its training objective with the CP validity requirements. Note that $\mathcal{L}_{\text{Validity}}$ is equivalent to the α-insensitivity loss used in robust SVMs (Vapnik et al. (1996)).

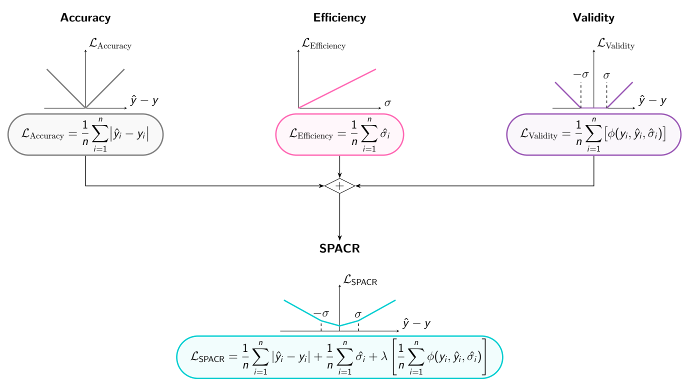


Figure 1: Decomposition of the SPACR loss. Top: the three functions ($\mathcal{L}_{\text{Accuracy}}$, $\mathcal{L}_{\text{Efficiency}}$, and $\mathcal{L}_{\text{Validity}}$). Center: their summation node. Bottom: the resulting overall SPACR loss.

Similar to other recent conformal training methods such as DOICR and ConfTr, SPACR relies on a post-hoc ICP calibration step to enforce distribution-free validity. Its contribution lies in learning better uncertainty estimates during training that align with conformal objectives. Although SPACR requires a differentiable model and thus departs from the model-agnostic spirit of classical CP, this assumption enables efficient, end-to-end optimization and supports dynamic confidence levels at inference without retraining. This trade-off aligns with practical needs in deep learning, where differentiability is standard.

Algorithm 1 outlines the SPACR training procedure. A key distinction of SPACR is that it does not require splitting the proper training set $\mathcal{D}_{train}$ into a primary training set and an embedded calibration set. This design allows us to be more sample-efficient when batch processing.

During inference, SPACR follows the standard post-hoc ICP calibration procedure, identical to traditional conformal methods and training-based techniques such as DOICR. Given a chosen significance level α and a separate calibration set $\mathcal{D}_{cal} = \{(x_j, y_j)\}_{j=1}^{n_{cal}}$, the model outputs predictions $\hat{y}_j$ and uncertainty estimates $\hat{\sigma}_j$. To prevent numerical instability for very small $\hat{\sigma}_j$, we compute the non-conformity scores as:

$$s_j = \frac{|y_j - \hat{y}_j|}{\hat{\sigma}_j + \beta} \quad \text{for } j = 1, \ldots, n_{cal}. \tag{6}$$

Algorithm 1 SPACR Training Procedure

Input: Proper training dataset $\mathcal{D}_{train}$, DNN $f_\theta = (f_\theta^{(1)}, f_\theta^{(2)})$, number of epochs $E_{\max}$, learning rate η
Output: Trained conformalized model f_θ

```
Initialize model parameters θ
for epoch = 1 to E_max do
    Shuffle and partition D_train into mini-batches 𝓑
    for each batch B ∈ 𝓑 do
        Forward pass: For each (x_i, y_i) ∈ B, compute predicted mean ŷ_i = f_θ^(1)(x_i) and uncertainty σ̂_i = exp(f_θ^(2)(x_i))
        Compute loss: Evaluate L_SPACR(B; θ) over the current batch B using Eq. 5
        Backward pass: Compute gradients ∇_θ L_SPACR
        Parameter update: Update θ using an optimizer (e.g., Adam) with learning rate η
    end for
end for
```

We set $\beta = 0.05\mu_{\hat{\sigma}}$ as a smoothing constant, where $\mu_{\hat{\sigma}}$ is the mean calibration uncertainty. For a fair comparison, this applies to all baselines using normalized non-conformity measures.

The conformity threshold $\hat{q}$ is set as the $\lceil (n_{cal}+1)(1-\alpha) \rceil$-th smallest value of these non-conformity scores. For a new input x_{test}, the final $(1-\alpha)$-CP interval is given by:

$$\hat{C}_{1-\alpha}(x_{test}) = \Big[\hat{y}_{test} - \hat{q}(\hat{\sigma}_{test} + \beta),\ \hat{y}_{test} + \hat{q}(\hat{\sigma}_{test} + \beta)\Big]. \tag{7}$$

Another key distinction of SPACR is that it does not require specifying a significance level α during training. The complexity of the calibration step remains identical to that of the standard ICP, but it crucially requires no model updates. This enables efficient exploration of multiple α values using a single trained model, which significantly reduces computational cost compared to methods that necessitate retraining for each desired α.

4 Experimental Setting

4.1 Methodology

To evaluate SPACR against existing approaches, we adopt a unified experimental setting where all methods share the same underlying neural network architecture, with differences tailored to each one:

- **Standard ICP (SICP)**: A conformal regression baseline trained via Mean Absolute Error (MAE) to predict $\hat{y}_i$, followed by post-hoc calibration to construct constant-width intervals (Eq. 1, Eq. 2).
- **Normalized ICP (NICP)**: Extends SICP by scaling uncertainties through absolute error-based difficulty estimation, leading to adaptive intervals (Eq. 3, Eq. 4).
- **Conformalized Quantile Regression (CQR)**: Romano et al. (2019) trains a model to predict lower and upper quantiles using the pinball loss then ensures validity with conformal calibration.
- **DOICR**: The Directly Optimized Inductive Conformal Regression method, as described in Subsection 2.2, which incorporates conformal calibration into training.
- **SPACR (Ours)**: The proposed Single-Pass Conformal Regressor (Section 3) integrates conformal goals directly into training, avoiding batch splitting and a fixed α.

For tabular data, all methods use a DNN with three 64-unit hidden layers. The output structure varies depending on the method: SICP and NICP use a single output neuron for point predictions, while DOICR and SPACR employ two output neurons, with one for the predicted mean and another for the uncertainty estimate. Additionally, for NICP, a separate detached layer is included to model the difficulty estimator. For the computer vision task, the DNN consists of fine-tuning a ResNet18 model pretrained on ImageNet1K.

All models are trained using the Adam optimizer with a learning rate of $\eta = 10^{-4}$, a batch size of 256, and a maximum of 200 epochs. For SPACR, the regularization parameter λ is set to 5. To ensure reproducibility, experiments are repeated across five independent random seeds. The framework is implemented in PyTorch.

4.2 Datasets

The experiments were conducted on various regression datasets, primarily tabular data and, to a lesser extent, image data. Each dataset was split into proper training, calibration, and testing sets with a ratio of 60%, 20%, 20%. The tabular datasets were sourced from the benchmark collection introduced by Grinsztajn et al. (2022) with different characteristics in terms of features and sample sizes. For the image-based datasets, we used Drift by Li et al. (2024), which contains satellite imagery for monitoring trees across five countries. Specifically, we selected 13,094 examples from Denmark to predict the number of trees per image. We also used the UTK Face dataset by Zhang et al. (2017), consisting of over 20,000 face images, to predict age.

4.3 Performance Metrics

To assess each method's effectiveness, we measure:

- **Coverage (%)**: The percentage of test samples for which the true target value falls within the predicted conformal interval. A well-calibrated method should ideally achieve a coverage rate close to the chosen confidence level $1 - \alpha$.
- **Efficiency**: The width of prediction intervals. Smaller interval widths indicate higher efficiency, meaning the method provides more precise uncertainty estimates. To characterize the distribution of interval widths, we use the median rather than the mean, as it is less sensitive to outliers.

5 Results

Due to space constraints, we present a subset of the experimental results in the main text and provide additional results in the Appendix A.

5.1 Performance (Coverage and Efficiency)

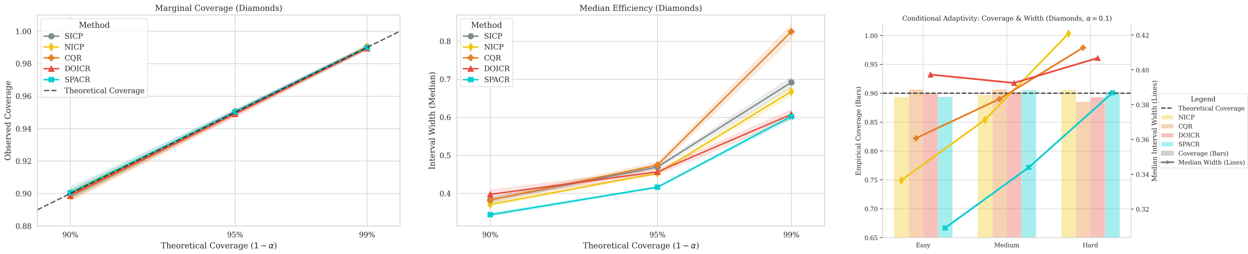
The experimental results comparing the marginal coverage rate and efficiency (median interval widths) of SICP, NICP, CQR, DOICR, and our proposed SPACR method across the different datasets at various chosen confidence levels ($\alpha \in \{0.10, 0.05, 0.01\}$) are presented in Table 1. Illustrative performance curves for the Diamonds and Drift datasets are shown in Figures 2 and 3, respectively.

Overall, the evaluated methods generally maintain marginal coverage within $\pm 1.5\%$ of the target confidence levels. As observed in Figures 2a and 3a, SPACR successfully satisfies the validity constraint, yielding coverage rates that track the nominal theoretical calibration line perfectly. A notable exception to this general robustness is DOICR, which exhibits severe under-coverage on certain datasets. For example, Table 1 shows DOICR achieving only $71.91 \pm 40.21\%$ for a target validity of 90% on Brazilian Houses, and severely failing on Drift with $57.04 \pm 52.08\%$ at 95%. This failure on Drift is further illustrated in Figure 3a, where the DOICR coverage curve deviates drastically from the theoretical baseline.

While providing valid marginal coverage, SPACR consistently delivers superior efficiency, giving the tightest prediction intervals in the majority of test cases. Across the 14 evaluated datasets in Table 1, SPACR achieves the lowest median interval widths, particularly at $\alpha = 0.10$ and $\alpha = 0.05$ levels. This advantage is clearly visualized in Figure 2b for Diamonds and Figure 3b for Drift, where the SPACR curve consistently lies below the competing baselines. The primary exceptions sometimes occur at the 99% confidence level, where competing baselines achieve slightly tighter intervals on a small subset of the datasets (e.g., DOICR on California, Medical Charges, and Superconduct), though SPACR remains highly competitive.

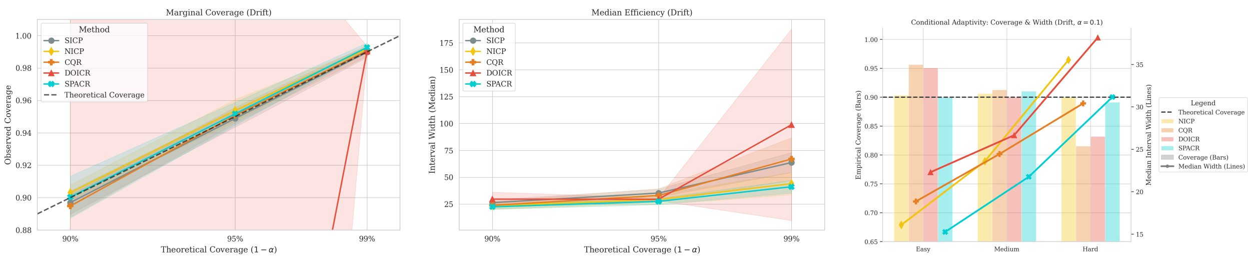
Table 1: Comparison of methods across datasets for Marginal Coverage and Median Efficiency (Mean ± Std). Best results are highlighted in bold, invalid coverage is marked in red.

Dataset	Method	$\alpha = 0.10$ (90%)		$\alpha = 0.05$ (95%)		$\alpha = 0.01$ (99%)	
		Cov. (%)	Med. Eff. ↓	Cov. (%)	Med. Eff. ↓	Cov. (%)	Med. Eff. ↓
Bike Sharing	SICP	90.13 ± 0.50	1.36 ± 0.02	94.79 ± 0.38	1.68 ± 0.02	99.13 ± 0.24	2.57 ± 0.09
	NICP	90.09 ± 0.14	**1.29 ± 0.04**	94.89 ± 0.49	1.55 ± 0.05	99.10 ± 0.27	2.20 ± 0.07
	CQR	89.76 ± 0.22	1.39 ± 0.06	94.69 ± 0.30	1.74 ± 0.05	**99.00 ± 0.35**	3.09 ± 0.08
	DOICR	90.07 ± 0.48	2.48 ± 0.22	95.20 ± 0.34	2.68 ± 0.22	99.03 ± 0.38	3.30 ± 0.23
	SPACR	**90.06 ± 0.52**	1.29 ± 0.09	**95.09 ± 0.41**	**1.54 ± 0.11**	99.02 ± 0.22	**2.07 ± 0.09**
Brazilian Houses	SICP	89.93 ± 0.52	**0.46 ± 0.04**	95.15 ± 0.28	**0.59 ± 0.04**	99.21 ± 0.16	0.94 ± 0.08
	NICP	89.72 ± 0.29	0.45 ± 0.05	94.98 ± 0.32	0.59 ± 0.06	99.23 ± 0.19	**0.93 ± 0.08**
	CQR	**89.97 ± 0.46**	0.51 ± 0.07	94.60 ± 0.73	0.66 ± 0.08	99.23 ± 0.26	1.23 ± 0.15
	DOICR	71.91 ± 40.21	1.53 ± 0.29	75.89 ± 42.42	1.87 ± 0.12	**99.14 ± 0.12**	2.47 ± 0.70
	SPACR	89.86 ± 0.69	1.07 ± 0.12	**95.00 ± 0.88**	1.22 ± 0.15	99.21 ± 0.22	1.59 ± 0.18
California	SICP	90.41 ± 0.80	0.28 ± 0.01	95.16 ± 0.62	0.36 ± 0.01	**99.01 ± 0.15**	0.56 ± 0.01
	NICP	90.34 ± 0.79	0.28 ± 0.01	95.08 ± 0.60	0.35 ± 0.01	99.01 ± 0.20	0.56 ± 0.02
	CQR	90.14 ± 0.92	0.30 ± 0.01	95.10 ± 0.36	0.38 ± 0.01	99.08 ± 0.17	0.61 ± 0.01
	DOICR	90.18 ± 1.16	0.28 ± 0.02	95.22 ± 0.56	0.34 ± 0.01	99.02 ± 0.06	**0.47 ± 0.01**
	SPACR	**90.06 ± 0.94**	**0.26 ± 0.01**	**95.04 ± 0.57**	**0.33 ± 0.01**	99.02 ± 0.15	0.53 ± 0.03
CPU Act	SICP	89.75 ± 1.37	**0.15 ± 0.01**	**94.97 ± 1.22**	**0.21 ± 0.02**	99.11 ± 0.31	0.58 ± 0.06
	NICP	89.79 ± 1.36	0.15 ± 0.01	95.25 ± 0.68	0.22 ± 0.02	99.12 ± 0.23	0.66 ± 0.17
	CQR	89.87 ± 0.94	0.16 ± 0.01	95.06 ± 1.18	0.28 ± 0.04	99.28 ± 0.26	0.78 ± 0.11
	DOICR	90.30 ± 0.47	0.97 ± 0.09	95.25 ± 0.76	1.12 ± 0.08	**99.02 ± 0.31**	1.28 ± 0.15
	SPACR	**89.96 ± 0.52**	0.24 ± 0.05	94.78 ± 0.71	0.27 ± 0.05	99.13 ± 0.15	**0.37 ± 0.05**
Diamonds	SICP	**89.99 ± 0.26**	0.38 ± 0.01	95.06 ± 0.15	0.47 ± 0.00	99.05 ± 0.11	0.69 ± 0.01
	NICP	89.89 ± 0.36	0.37 ± 0.01	94.91 ± 0.05	0.45 ± 0.00	99.03 ± 0.10	0.67 ± 0.01
	CQR	89.94 ± 0.23	0.38 ± 0.01	94.91 ± 0.17	0.47 ± 0.01	98.98 ± 0.10	0.83 ± 0.02
	DOICR	89.86 ± 0.27	0.40 ± 0.01	94.91 ± 0.16	0.46 ± 0.01	98.94 ± 0.10	0.61 ± 0.01
	SPACR	90.05 ± 0.36	**0.34 ± 0.00**	**95.02 ± 0.11**	**0.42 ± 0.00**	**99.00 ± 0.08**	**0.60 ± 0.01**
Fifa	SICP	90.22 ± 0.59	0.30 ± 0.01	95.15 ± 0.52	0.43 ± 0.01	99.08 ± 0.15	0.65 ± 0.02
	NICP	90.11 ± 0.46	0.30 ± 0.01	95.23 ± 0.47	0.44 ± 0.01	**99.04 ± 0.29**	0.63 ± 0.02
	CQR	**90.07 ± 0.70**	0.33 ± 0.00	94.86 ± 0.52	0.36 ± 0.00	99.21 ± 0.27	**0.43 ± 0.01**
	DOICR	90.10 ± 0.47	0.32 ± 0.01	**94.98 ± 0.52**	0.38 ± 0.01	99.08 ± 0.28	0.44 ± 0.02
	SPACR	90.27 ± 0.51	**0.28 ± 0.00**	94.90 ± 0.57	**0.36 ± 0.01**	98.93 ± 0.16	0.61 ± 0.04
House Sales	SICP	**89.87 ± 0.74**	0.56 ± 0.01	94.76 ± 0.39	0.72 ± 0.02	98.85 ± 0.32	1.13 ± 0.06
	NICP	89.65 ± 0.65	**0.54 ± 0.02**	94.53 ± 0.51	**0.68 ± 0.02**	98.84 ± 0.31	1.04 ± 0.03
	CQR	89.63 ± 0.85	0.54 ± 0.03	94.66 ± 0.44	0.69 ± 0.03	98.80 ± 0.36	1.13 ± 0.05
	DOICR	90.28 ± 0.45	2.30 ± 0.28	**94.99 ± 0.32**	2.56 ± 0.37	98.84 ± 0.08	2.64 ± 0.28
	SPACR	89.83 ± 1.04	0.55 ± 0.02	94.74 ± 0.41	0.68 ± 0.03	**98.87 ± 0.31**	**1.00 ± 0.03**
Isolet	SICP	90.29 ± 1.69	0.88 ± 0.03	94.73 ± 1.45	1.19 ± 0.08	99.40 ± 0.12	2.33 ± 0.22
	NICP	89.71 ± 1.49	0.87 ± 0.03	**95.06 ± 1.60**	1.20 ± 0.11	99.33 ± 0.11	2.23 ± 0.17
	CQR	89.68 ± 1.51	0.87 ± 0.03	94.69 ± 1.47	1.15 ± 0.03	**99.00 ± 0.18**	2.01 ± 0.04
	DOICR	**90.23 ± 0.89**	1.13 ± 0.04	94.82 ± 0.41	1.29 ± 0.04	99.03 ± 0.28	1.70 ± 0.09
	SPACR	89.05 ± 1.26	**0.78 ± 0.06**	94.54 ± 0.70	**1.01 ± 0.07**	99.13 ± 0.28	**1.61 ± 0.11**
Medical Charges	SICP	89.85 ± 0.19	0.18 ± 0.00	94.91 ± 0.22	0.27 ± 0.01	98.99 ± 0.12	0.59 ± 0.02
	NICP	89.95 ± 0.15	0.18 ± 0.00	94.93 ± 0.19	0.26 ± 0.01	99.04 ± 0.09	0.55 ± 0.02
	CQR	**89.96 ± 0.08**	0.19 ± 0.01	94.96 ± 0.17	0.29 ± 0.01	98.98 ± 0.11	0.80 ± 0.04
	DOICR	90.12 ± 0.17	0.21 ± 0.01	94.92 ± 0.25	0.26 ± 0.01	99.00 ± 0.05	**0.37 ± 0.01**
	SPACR	89.90 ± 0.09	**0.17 ± 0.00**	**95.00 ± 0.20**	**0.24 ± 0.00**	**99.00 ± 0.07**	0.50 ± 0.01
Pol	SICP	90.07 ± 0.80	0.35 ± 0.02	95.12 ± 0.77	0.94 ± 0.08	98.97 ± 0.15	3.44 ± 0.34
	NICP	**90.06 ± 0.83**	0.34 ± 0.02	95.11 ± 0.71	0.91 ± 0.07	99.03 ± 0.13	3.33 ± 0.27
	CQR	90.28 ± 0.68	0.49 ± 0.08	95.27 ± 0.23	1.32 ± 0.19	**98.99 ± 0.16**	2.48 ± 0.14
	DOICR	89.39 ± 0.87	0.79 ± 0.15	**94.92 ± 0.19**	1.20 ± 0.12	98.95 ± 0.32	1.72 ± 0.14
	SPACR	90.37 ± 0.43	**0.09 ± 0.03**	95.67 ± 0.37	**0.14 ± 0.04**	99.14 ± 0.17	**0.33 ± 0.08**
Superconduct	SICP	89.83 ± 0.56	1.66 ± 0.04	94.88 ± 0.27	2.24 ± 0.06	**98.96 ± 0.28**	3.65 ± 0.18
	NICP	**89.92 ± 0.61**	1.52 ± 0.07	94.86 ± 0.44	2.06 ± 0.09	98.94 ± 0.24	3.50 ± 0.15
	CQR	89.56 ± 0.47	1.59 ± 0.02	94.59 ± 0.23	2.02 ± 0.02	98.93 ± 0.38	3.26 ± 0.11
	DOICR	89.74 ± 0.61	1.55 ± 0.05	94.85 ± 0.34	1.85 ± 0.04	98.91 ± 0.30	**2.60 ± 0.09**
	SPACR	89.79 ± 0.36	**1.29 ± 0.04**	**95.02 ± 0.27**	**1.69 ± 0.05**	98.83 ± 0.16	3.03 ± 0.11
Wine Quality	SICP	90.46 ± 1.03	**0.34 ± 0.01**	**94.95 ± 0.64**	0.43 ± 0.02	**99.00 ± 0.37**	0.65 ± 0.06
	NICP	91.31 ± 1.22	0.36 ± 0.02	95.32 ± 0.89	0.44 ± 0.03	98.92 ± 0.26	0.67 ± 0.05
	CQR	90.26 ± 1.06	0.35 ± 0.01	94.80 ± 0.70	0.46 ± 0.02	99.11 ± 0.46	0.77 ± 0.05
	DOICR	**90.02 ± 0.75**	0.43 ± 0.04	95.46 ± 0.77	0.50 ± 0.04	99.18 ± 0.20	0.68 ± 0.04
	SPACR	90.48 ± 0.85	0.35 ± 0.01	95.31 ± 0.80	**0.43 ± 0.01**	99.02 ± 0.33	**0.65 ± 0.05**
Drift	SICP	89.70 ± 0.71	26.41 ± 2.73	94.89 ± 0.54	35.33 ± 3.64	**98.97 ± 0.32**	63.70 ± 9.28
	NICP	90.31 ± 0.47	23.55 ± 3.32	95.38 ± 0.67	29.35 ± 4.74	99.18 ± 0.11	44.13 ± 10.32
	CQR	89.48 ± 0.69	23.94 ± 3.59	**95.08 ± 0.49**	33.06 ± 6.29	99.06 ± 0.33	66.99 ± 19.81
	DOICR	53.65 ± 48.98	29.70 ± 6.46	57.04 ± 52.08	29.56 ± 1.34	99.04 ± 0.30	98.84 ± 89.02
	SPACR	**90.03 ± 1.29**	**22.51 ± 2.28**	95.17 ± 0.71	**27.59 ± 2.98**	99.27 ± 0.28	**41.19 ± 5.90**
UTK Face	SICP	89.97 ± 0.67	32.15 ± 4.12	94.98 ± 0.45	40.67 ± 5.42	98.99 ± 0.23	60.17 ± 6.94
	NICP	90.16 ± 0.35	30.36 ± 9.72	94.92 ± 0.34	38.97 ± 13.57	99.08 ± 0.03	58.48 ± 16.74
	CQR	**90.01 ± 0.33**	26.29 ± 3.96	95.01 ± 0.43	41.82 ± 8.57	**99.00 ± 0.14**	55.22 ± 6.05
	DOICR	89.46 ± 0.29	31.67 ± 2.57	**95.00 ± 0.22**	38.38 ± 6.59	98.94 ± 0.25	47.42 ± 5.11
	SPACR	90.13 ± 0.46	**21.84 ± 0.84**	95.02 ± 0.30	**27.49 ± 0.98**	98.95 ± 0.14	**42.51 ± 2.06**



(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 2: Performance figures for the different approaches for the Diamonds dataset.



(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 3: Performance figures for the different approaches for the Drift dataset.

5.2 Adaptability of SPACR

The empirical results highlight SPACR's robust conditional adaptivity across varying data characteristics and user requirements. By decoupling training from the choice of α, SPACR adapts seamlessly to different confidence requirements where a single trained model produces valid intervals at 90%, 95%, and 99% coverage (Table 1) without the extra training computation costs required by other methods such as CQR and DOICR.

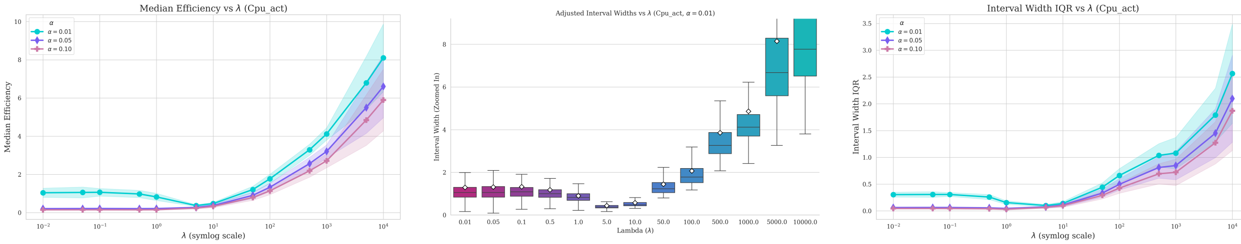
Figures 2c and 3c highlight SPACR's conditional adaptivity across difficulty bins ("Easy", "Medium", "Hard") at $\alpha = 0.10$. Although exact distribution-free conditional coverage is theoretically impossible without strong assumptions Foygel Barber et al. (2021), SPACR's optimization serves as a highly effective heuristic, maintaining empirical coverage near the target across subsets. In contrast, baselines struggle with consistent conditional calibration. For instance, on the Drift dataset (Figure 3c), CQR and DOICR suffer from severe under-coverage in the "Hard" bin, which is compensated by an over-coverage in the "Easy" bin.

Furthermore, SPACR achieves this stable coverage while maintaining highly adaptive and efficient prediction intervals. As shown by the median interval width lines in Figures 2c and 3c, SPACR's interval width scales smoothly and monotonically in response to instance difficulty, widening appropriately for "Hard" samples while remaining significantly tighter than the other methods across all bins. For instance, in Figure 2c, DOICR exhibits a nearly flat median interval width line across the difficulty bins. This lack of an upward trajectory indicates that DOICR fails to scale its uncertainty estimates according to instance difficulty, essentially outputting rigid, unadaptive intervals. These observations confirm that SPACR's learned $\hat{\sigma}$ successfully adapts to the residual distributions without requiring a predefined α parameter during training.

In conclusion, SPACR flexibly adjusts to diverse data settings (from low-dimensional tabular benchmarks to high-dimensional image tasks) and dynamically optimizes the trade-off between efficiency and coverage, making it a versatile tool for uncertainty quantification in supervised regression tasks.

5.3 Sensitivity Analysis of the Regularization Parameter λ

The regularization parameter λ in the $\mathcal{L}_{\text{SPACR}}$ objective governs the trade-off between point prediction accuracy ($\mathcal{L}_{\text{Accuracy}}$), interval efficiency ($\mathcal{L}_{\text{Efficiency}}$), and empirical coverage validity ($\mathcal{L}_{\text{Validity}}$). By modeling



(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 4: Results analysis for varying λ values for the CPU Act dataset.

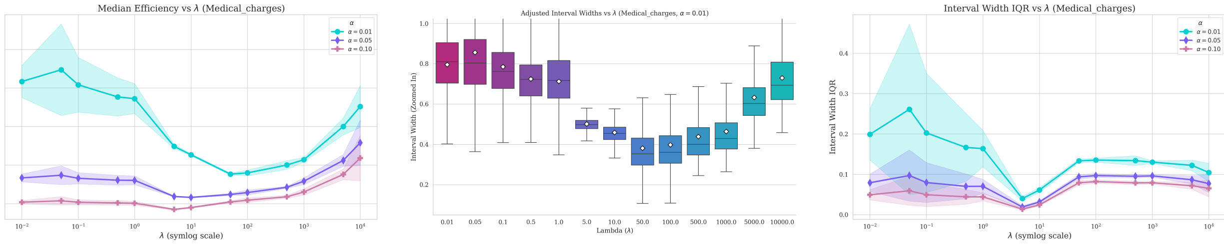
this trade-off with a single hyperparameter, we enable a highly efficient 1D logarithmic search, completely avoiding the exponential scaling and computational burden associated with multi-parameter grid searches. To evaluate hyperparameter sensitivity, performance is analyzed across a logarithmic grid from $\lambda = 0.01$ to 10000 and reported in Table 2, supplemented by the visual analysis of the width boxplots and efficiency curves for the CPU Act (Figure 4), Medical Charges (Figure 5), and UTK Face (Figure 6) datasets.

A primary observation is that the marginal coverage remains stable near the nominal $1 - \alpha$ target across all λ values, confirming that post-hoc ICP guarantees marginal validity independent of the model's uncertainty estimates. However, the choice of λ significantly impacts interval efficiency and conditional adaptivity:

- **Under-regularization ($\lambda \to 0$):** The validity term $\mathcal{L}_{\text{Validity}}$ becomes negligible, driving uncertainty estimates toward zero. This manifests in two ways depending on the dataset's complexity. On datasets like CPU Act, the network uniformly shrinks $\hat{\sigma}_i$ to a tiny constant, which yields non-adaptive intervals resembling Standard ICP. For example, for $\alpha = 0.10$ and $\lambda = 0.01$, IQR $= 0.04 \approx 0$ as seen in Figure 4c. On datasets like UTK Face, driving $\hat{\sigma}_i \to 0$ introduces vanishing denominators in the non-conformity scores, which inflates them and forces massive calibration quantiles $\hat{q}$, resulting in impractically wide, high-variance intervals. For instance, at $\alpha = 0.10$, setting $\lambda = 0.01$ yields a highly unstable median efficiency of 138.69 and an IQR of 235.76 (Figure 6c).
- **Over-regularization ($\lambda \to \infty$):** The validity term $\mathcal{L}_{\text{Validity}}$ dominates, forcing the model to prioritize satisfying the coverage constraint $\phi(y_i, \hat{y}_i, \hat{\sigma}_i)$ regardless of interval width or point-prediction error. As a result, non-conformity scores approach zero, and the final intervals are overwhelmingly dictated by the inflated $\hat{\sigma}_i$.63 The visual manifestation of this degradation bifurcates into two distinct behaviors depending on the optimization dynamics of the specific dataset. In the first case, observed in datasets like UTK Face (Figure 6), the median efficiency degrades significantly (e.g., from 21.84 at $\lambda = 5$ to 76.74 at $\lambda = 10000$ for $\alpha = 0.10$), but the IQR remains artificially compressed (hovering around 12.56). In the second case, observed in datasets like CPU Act (Figure 4) and Medical Charges (Figure 5), both the median interval width and the variance of those widths explode (e.g., at $\alpha = 0.10$ for CPU Act, increasing λ from 5 to 10000 degrades the median efficiency from 0.24 with an IQR of 0.06 to 5.89 with an IQR of 1.61).
- **Optimal Balance ($\lambda \in [1, 50]$):** Moderate values ensure valid coverage without unnecessary interval inflation. Across the datasets, setting $\lambda = 5$ consistently yields near-optimal median efficiency while maintaining strong conditional adaptivity (a healthy, non-zero IQR). Moreover, it is even possible to optimize results by choosing a suitable λ. For instance, by checking the efficiency curves in Figure 5, it is possible to determine that choosing $\lambda = 50$ when $\alpha = 0.01$ as in Table 2 gives better efficiency results compared to DOICR in Table 1.

5.4 Ablation Study: Impact of Validation and Efficiency Loss Terms

To evaluate the individual contributions of the components in our proposed loss function, we conducted an ablation study. We compare the full SPACR model against two variants: **SPACR_VAL**, which ablates the

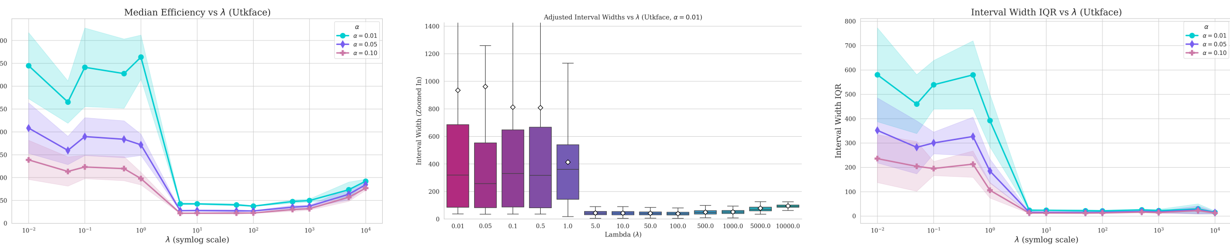


(a) Median Efficiency.

(b) Width Boxplots.

(c) Interquartile Range (IQR).

Figure 5: Results analysis for varying λ values for the Medical Charges dataset.



(a) Median Efficiency.

(b) Width Boxplots.

(c) Interquartile Range (IQR).

Figure 6: Results analysis for varying λ values for the UTK Face dataset.

Table 2: Impact of λ on Marginal Coverage, Median Efficiency, and Interval Width IQR. Best results per dataset and confidence level are highlighted in bold. The row corresponding to $\lambda = 5$ is shaded in gray.

Dataset	**Lambda (λ)**	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
		Cov. (%)	Med. Eff.	IQR	Cov. (%)	Med. Eff.	IQR	Cov. (%)	Med. Eff.	IQR
	0.01	89.59 ± 1.36	**0.15 ± 0.01**	0.04 ± 0.01	94.74 ± 1.16	**0.21 ± 0.02**	0.06 ± 0.01	98.97 ± 0.30	1.04 ± 0.24	0.30 ± 0.05
	0.05	89.49 ± 1.07	**0.15 ± 0.01**	0.04 ± 0.01	94.80 ± 1.30	**0.21 ± 0.02**	0.06 ± 0.01	99.02 ± 0.26	1.06 ± 0.29	0.31 ± 0.05
	0.1	89.46 ± 1.23	**0.15 ± 0.01**	0.04 ± 0.01	94.70 ± 1.26	0.22 ± 0.02	0.06 ± 0.01	**99.02 ± 0.22**	1.06 ± 0.20	0.31 ± 0.04
	0.5	89.05 ± 0.78	**0.15 ± 0.01**	0.04 ± 0.01	94.77 ± 0.84	**0.21 ± 0.02**	0.06 ± 0.01	98.94 ± 0.11	0.98 ± 0.19	0.26 ± 0.05
	1	88.82 ± 1.03	0.15 ± 0.02	**0.03 ± 0.02**	**94.94 ± 0.85**	**0.21 ± 0.02**	**0.04 ± 0.02**	99.05 ± 0.16	0.82 ± 0.19	0.16 ± 0.03
	5	89.96 ± 0.52	0.24 ± 0.05	0.06 ± 0.02	94.78 ± 0.71	0.27 ± 0.05	0.07 ± 0.02	99.13 ± 0.15	**0.37 ± 0.05**	**0.10 ± 0.02**
CPU Act	10	**90.01 ± 0.59**	0.33 ± 0.07	0.10 ± 0.03	95.14 ± 0.50	0.37 ± 0.07	0.11 ± 0.03	99.08 ± 0.26	0.47 ± 0.06	0.14 ± 0.03
	50	90.01 ± 0.62	0.79 ± 0.15	0.29 ± 0.07	94.81 ± 0.48	0.91 ± 0.17	0.33 ± 0.08	99.15 ± 0.54	1.21 ± 0.18	0.44 ± 0.08
	100	90.26 ± 1.20	1.14 ± 0.20	0.43 ± 0.10	95.08 ± 0.63	1.33 ± 0.22	0.50 ± 0.11	99.15 ± 0.40	1.77 ± 0.24	0.66 ± 0.13
	500	90.27 ± 0.85	2.19 ± 0.33	0.69 ± 0.19	95.23 ± 0.54	2.57 ± 0.33	0.81 ± 0.20	99.21 ± 0.34	3.29 ± 0.29	1.04 ± 0.22
	1000	90.13 ± 0.63	2.71 ± 0.37	0.72 ± 0.25	95.24 ± 0.53	3.20 ± 0.34	0.85 ± 0.27	99.23 ± 0.29	4.12 ± 0.32	1.08 ± 0.29
	5000	90.19 ± 0.83	4.84 ± 1.33	1.27 ± 0.39	94.91 ± 0.23	5.50 ± 1.35	1.45 ± 0.46	99.10 ± 0.22	6.79 ± 1.39	1.79 ± 0.50
	10000	90.23 ± 0.71	5.89 ± 1.61	1.87 ± 0.74	95.23 ± 0.52	6.60 ± 1.64	2.10 ± 0.81	99.06 ± 0.31	8.10 ± 1.77	2.56 ± 0.93
	0.01	90.01 ± 0.14	0.21 ± 0.01	0.05 ± 0.01	94.97 ± 0.28	0.33 ± 0.02	0.08 ± 0.02	98.96 ± 0.09	0.83 ± 0.08	0.20 ± 0.06
	0.05	90.02 ± 0.20	0.21 ± 0.02	0.06 ± 0.04	94.96 ± 0.27	0.35 ± 0.05	0.10 ± 0.06	98.96 ± 0.09	0.89 ± 0.24	0.26 ± 0.21
	0.1	89.97 ± 0.23	0.21 ± 0.01	0.05 ± 0.03	94.91 ± 0.28	0.33 ± 0.03	0.08 ± 0.05	98.95 ± 0.07	0.82 ± 0.14	0.20 ± 0.15
	0.5	89.90 ± 0.15	0.20 ± 0.01	0.04 ± 0.02	94.95 ± 0.24	0.32 ± 0.03	0.07 ± 0.03	98.97 ± 0.09	0.75 ± 0.10	0.17 ± 0.08
	1	90.01 ± 0.14	0.20 ± 0.01	0.04 ± 0.01	94.99 ± 0.19	0.32 ± 0.02	0.07 ± 0.02	98.98 ± 0.06	0.74 ± 0.08	0.16 ± 0.05
	5	89.90 ± 0.09	**0.17 ± 0.00**	**0.01 ± 0.00**	**95.00 ± 0.20**	0.24 ± 0.00	**0.02 ± 0.00**	**99.00 ± 0.07**	0.50 ± 0.01	**0.04 ± 0.01**
Medical Charges	10	90.03 ± 0.14	0.18 ± 0.00	0.02 ± 0.00	94.98 ± 0.13	**0.23 ± 0.00**	0.03 ± 0.00	99.03 ± 0.05	0.45 ± 0.01	0.06 ± 0.01
	50	**90.00 ± 0.14**	0.21 ± 0.01	0.08 ± 0.01	95.05 ± 0.12	0.25 ± 0.01	0.09 ± 0.01	99.07 ± 0.04	**0.35 ± 0.01**	0.13 ± 0.00
	100	90.11 ± 0.11	0.22 ± 0.01	0.08 ± 0.00	95.04 ± 0.08	0.26 ± 0.01	0.10 ± 0.00	99.02 ± 0.08	0.36 ± 0.01	0.14 ± 0.00
	500	90.08 ± 0.27	0.24 ± 0.01	0.08 ± 0.00	95.04 ± 0.08	0.29 ± 0.00	0.10 ± 0.00	99.00 ± 0.04	0.40 ± 0.02	0.13 ± 0.01
	1000	89.97 ± 0.18	0.26 ± 0.02	0.08 ± 0.00	94.94 ± 0.13	0.32 ± 0.02	0.10 ± 0.00	99.01 ± 0.04	0.43 ± 0.01	0.13 ± 0.00
	5000	90.02 ± 0.37	0.35 ± 0.03	0.07 ± 0.01	95.11 ± 0.20	0.42 ± 0.03	0.09 ± 0.01	99.00 ± 0.10	0.60 ± 0.04	0.12 ± 0.01
	10000	89.98 ± 0.22	0.44 ± 0.12	0.07 ± 0.02	95.10 ± 0.22	0.52 ± 0.12	0.08 ± 0.02	99.01 ± 0.05	0.70 ± 0.11	0.10 ± 0.02
	0.01	90.02 ± 0.65	138.69 ± 42.87	235.76 ± 98.26	95.06 ± 0.42	208.32 ± 55.31	351.79 ± 133.48	99.01 ± 0.19	344.87 ± 72.34	580.14 ± 192.72
	0.05	90.19 ± 0.72	113.36 ± 32.09	204.10 ± 102.21	94.85 ± 0.55	159.39 ± 30.99	282.22 ± 108.78	99.02 ± 0.27	265.22 ± 46.35	459.76 ± 120.96
	0.1	90.13 ± 0.48	123.09 ± 25.53	195.18 ± 28.97	94.78 ± 0.54	189.65 ± 41.42	300.17 ± 44.72	98.86 ± 0.19	341.30 ± 86.09	538.98 ± 99.85
	0.5	89.79 ± 0.46	119.50 ± 26.40	213.17 ± 54.07	94.71 ± 0.39	183.79 ± 40.42	326.68 ± 79.70	98.88 ± 0.22	327.50 ± 75.78	579.72 ± 139.87
	1	89.68 ± 0.76	97.68 ± 13.83	106.05 ± 31.99	94.52 ± 0.56	171.63 ± 23.31	184.91 ± 49.37	98.81 ± 0.22	363.66 ± 48.17	392.12 ± 108.11
	5	90.13 ± 0.46	**21.84 ± 0.84**	**12.18 ± 1.61**	95.02 ± 0.30	27.49 ± 0.98	**15.36 ± 2.26**	98.95 ± 0.14	42.51 ± 2.06	23.82 ± 4.13
UTK Face	10	90.02 ± 0.20	21.96 ± 1.16	12.34 ± 1.18	95.09 ± 0.31	27.79 ± 1.66	15.61 ± 1.39	99.01 ± 0.25	42.41 ± 1.56	23.80 ± 1.50
	50	90.66 ± 0.85	22.20 ± 2.17	11.97 ± 2.00	94.97 ± 0.45	27.17 ± 2.55	14.67 ± 2.55	99.03 ± 0.06	40.26 ± 2.33	21.88 ± 4.48
	100	89.81 ± 0.38	22.38 ± 0.66	12.65 ± 1.90	95.07 ± 0.18	**26.87 ± 0.79**	15.18 ± 2.18	98.92 ± 0.17	**37.44 ± 0.73**	21.17 ± 3.23
	500	90.30 ± 0.34	29.87 ± 3.94	15.93 ± 2.13	95.39 ± 0.56	35.25 ± 4.07	18.82 ± 2.40	99.08 ± 0.11	47.47 ± 3.85	25.38 ± 2.95
	1000	89.73 ± 0.64	31.78 ± 2.96	14.13 ± 3.13	94.85 ± 0.38	37.36 ± 2.49	16.69 ± 4.00	98.95 ± 0.21	49.66 ± 3.48	22.27 ± 5.78
	5000	90.08 ± 0.35	56.83 ± 9.02	22.25 ± 13.65	**94.98 ± 0.15**	63.41 ± 12.67	25.16 ± 16.82	**98.99 ± 0.22**	73.31 ± 16.82	29.33 ± 20.77
	10000	**90.00 ± 0.34**	76.74 ± 4.66	12.56 ± 4.47	95.03 ± 0.45	84.29 ± 4.72	13.83 ± 5.02	98.91 ± 0.24	92.08 ± 4.38	**15.22 ± 5.88**

efficiency penalty ($\mathcal{L}_{Efficiency}$), and **SPACR_EFF**, which ablates the validity penalty ($\mathcal{L}_{Validity}$). Table 3 reports the marginal coverage, median efficiency, and the interquartile range (IQR) of the interval widths across the evaluated datasets for $\alpha \in \{0.10, 0.05, 0.01\}$.

Table 3: Ablation Study: Impact of Validation and Efficiency loss terms.

Dataset	Variant	$\alpha = 0.10$ (90%)			$\alpha = 0.05$ (95%)			$\alpha = 0.01$ (99%)		
		Cov. (%)	Med. Eff. ↓	IQR ↓	Cov. (%)	Med. Eff. ↓	IQR ↓	Cov. (%)	Med. Eff. ↓	IQR ↓
Bike Sharing	SPACR_VAL	89.96 ± 0.69	1.59 ± 0.12	0.44 ± 0.18	94.95 ± 0.42	2.02 ± 0.17	0.56 ± 0.24	99.00 ± 0.21	3.19 ± 0.31	0.88 ± 0.39
	SPACR_EFF	90.23 ± 0.20	1.58 ± 0.15	1.98 ± 0.14	95.51 ± 0.53	2.23 ± 0.30	2.79 ± 0.23	99.09 ± 0.14	3.96 ± 1.25	4.87 ± 1.01
	SPACR	90.06 ± 0.52	1.29 ± 0.09	0.49 ± 0.06	95.09 ± 0.41	1.54 ± 0.11	0.58 ± 0.07	99.02 ± 0.22	2.07 ± 0.09	0.79 ± 0.12
Brazilian Houses	SPACR_VAL	90.40 ± 0.46	0.71 ± 0.10	0.27 ± 0.16	95.32 ± 0.16	0.86 ± 0.12	0.33 ± 0.20	99.13 ± 0.26	1.25 ± 0.15	0.48 ± 0.28
	SPACR_EFF	89.92 ± 0.86	0.73 ± 0.19	0.59 ± 0.18	95.16 ± 0.64	0.99 ± 0.29	0.81 ± 0.28	99.08 ± 0.24	1.68 ± 0.47	1.38 ± 0.48
	SPACR	89.86 ± 0.69	1.07 ± 0.12	0.08 ± 0.01	95.00 ± 0.88	1.22 ± 0.15	0.09 ± 0.01	99.21 ± 0.22	1.59 ± 0.18	0.12 ± 0.02
California	SPACR_VAL	90.40 ± 0.80	0.28 ± 0.01	0.00 ± 0.00	95.24 ± 0.62	0.36 ± 0.01	0.00 ± 0.00	99.03 ± 0.16	0.56 ± 0.01	0.00 ± 0.00
	SPACR_EFF	90.47 ± 1.04	0.30 ± 0.01	0.06 ± 0.01	95.42 ± 0.75	0.39 ± 0.02	0.07 ± 0.02	99.11 ± 0.21	0.66 ± 0.03	0.12 ± 0.02
	SPACR	90.06 ± 0.94	0.26 ± 0.01	0.06 ± 0.00	95.04 ± 0.57	0.33 ± 0.01	0.07 ± 0.00	99.02 ± 0.15	0.53 ± 0.03	0.12 ± 0.01
CPU Act	SPACR_VAL	89.98 ± 1.18	0.23 ± 0.05	0.04 ± 0.01	94.95 ± 0.15	0.32 ± 0.07	0.06 ± 0.01	99.07 ± 0.38	1.33 ± 0.82	0.22 ± 0.10
	SPACR_EFF	89.47 ± 1.39	0.15 ± 0.01	0.04 ± 0.01	94.61 ± 1.32	0.21 ± 0.01	0.06 ± 0.01	99.04 ± 0.29	1.06 ± 0.28	0.31 ± 0.05
	SPACR	89.96 ± 0.52	0.24 ± 0.05	0.06 ± 0.02	94.78 ± 0.71	0.27 ± 0.05	0.07 ± 0.02	99.13 ± 0.15	0.37 ± 0.05	0.10 ± 0.02
Diamonds	SPACR_VAL	89.93 ± 0.28	0.38 ± 0.00	0.05 ± 0.01	95.02 ± 0.13	0.47 ± 0.00	0.06 ± 0.02	99.00 ± 0.11	0.66 ± 0.01	0.08 ± 0.02
	SPACR_EFF	89.89 ± 0.46	0.54 ± 0.02	0.62 ± 0.08	94.98 ± 0.31	0.84 ± 0.05	0.97 ± 0.13	99.09 ± 0.09	1.76 ± 0.07	2.03 ± 0.21
	SPACR	90.05 ± 0.36	0.34 ± 0.00	0.06 ± 0.00	95.02 ± 0.11	0.42 ± 0.00	0.07 ± 0.01	99.00 ± 0.08	0.60 ± 0.01	0.10 ± 0.01
Fifa	SPACR_VAL	90.15 ± 0.68	0.30 ± 0.01	0.02 ± 0.00	95.13 ± 0.63	0.43 ± 0.02	0.02 ± 0.01	99.06 ± 0.23	0.64 ± 0.02	0.04 ± 0.01
	SPACR_EFF	90.17 ± 0.59	0.34 ± 0.02	0.06 ± 0.02	95.23 ± 0.20	0.49 ± 0.02	0.08 ± 0.03	99.18 ± 0.08	0.97 ± 0.22	0.16 ± 0.08
	SPACR	90.27 ± 0.51	0.28 ± 0.00	0.05 ± 0.00	94.90 ± 0.57	0.36 ± 0.01	0.07 ± 0.01	98.93 ± 0.16	0.61 ± 0.04	0.12 ± 0.02
House Sales	SPACR_VAL	89.83 ± 0.91	0.57 ± 0.01	0.06 ± 0.01	94.79 ± 0.35	0.73 ± 0.01	0.08 ± 0.01	98.76 ± 0.24	1.12 ± 0.04	0.12 ± 0.02
	SPACR_EFF	89.72 ± 0.68	0.60 ± 0.02	0.20 ± 0.01	94.80 ± 0.37	0.79 ± 0.03	0.26 ± 0.01	98.92 ± 0.26	1.27 ± 0.06	0.42 ± 0.02
	SPACR	89.83 ± 1.04	0.55 ± 0.02	0.10 ± 0.01	94.74 ± 0.41	0.68 ± 0.03	0.12 ± 0.01	98.87 ± 0.31	1.00 ± 0.03	0.18 ± 0.01
Isolet	SPACR_VAL	89.82 ± 1.36	0.92 ± 0.02	0.27 ± 0.05	94.76 ± 1.26	1.26 ± 0.08	0.37 ± 0.07	99.23 ± 0.26	2.39 ± 0.22	0.70 ± 0.10
	SPACR_EFF	90.32 ± 0.93	1.38 ± 0.25	1.88 ± 0.52	95.27 ± 0.71	2.08 ± 0.32	2.82 ± 0.70	99.06 ± 0.35	4.04 ± 0.66	5.50 ± 1.43
	SPACR	89.05 ± 1.26	0.78 ± 0.06	0.43 ± 0.03	94.54 ± 0.70	1.01 ± 0.07	0.56 ± 0.05	99.13 ± 0.28	1.61 ± 0.11	0.90 ± 0.09
Medical Charges	SPACR_VAL	89.95 ± 0.18	0.18 ± 0.00	0.00 ± 0.00	94.93 ± 0.16	0.27 ± 0.01	0.01 ± 0.00	99.01 ± 0.10	0.58 ± 0.03	0.01 ± 0.01
	SPACR_EFF	90.03 ± 0.22	0.22 ± 0.02	0.06 ± 0.04	94.98 ± 0.22	0.35 ± 0.05	0.10 ± 0.06	98.96 ± 0.10	0.93 ± 0.29	0.28 ± 0.24
	SPACR	89.90 ± 0.09	0.17 ± 0.00	0.01 ± 0.00	95.00 ± 0.20	0.24 ± 0.00	0.02 ± 0.00	99.00 ± 0.07	0.50 ± 0.01	0.04 ± 0.01
Pol	SPACR_VAL	89.99 ± 0.83	0.53 ± 0.05	0.87 ± 0.26	95.05 ± 0.68	1.33 ± 0.08	2.19 ± 0.60	99.05 ± 0.27	4.89 ± 0.34	8.10 ± 2.32
	SPACR_EFF	90.07 ± 0.65	0.48 ± 0.04	0.63 ± 0.05	95.47 ± 0.64	0.84 ± 0.10	1.11 ± 0.07	99.07 ± 0.29	2.17 ± 0.25	2.88 ± 0.32
	SPACR	90.37 ± 0.43	0.09 ± 0.03	0.33 ± 0.03	95.67 ± 0.37	0.14 ± 0.04	0.51 ± 0.06	99.14 ± 0.17	0.33 ± 0.08	1.17 ± 0.08
Superconduct	SPACR_VAL	89.89 ± 0.39	2.00 ± 0.05	0.97 ± 0.12	94.88 ± 0.35	2.65 ± 0.10	1.29 ± 0.18	99.08 ± 0.20	4.16 ± 0.18	2.02 ± 0.30
	SPACR_EFF	90.40 ± 0.60	1.70 ± 0.06	2.45 ± 0.23	95.22 ± 0.61	2.69 ± 0.13	3.88 ± 0.47	99.09 ± 0.28	6.06 ± 0.43	8.68 ± 0.62
	SPACR	89.79 ± 0.36	1.29 ± 0.04	0.47 ± 0.02	95.02 ± 0.27	1.69 ± 0.05	0.62 ± 0.03	98.83 ± 0.16	3.03 ± 0.11	1.11 ± 0.06
Wine Quality	SPACR_VAL	90.45 ± 1.29	0.35 ± 0.01	0.05 ± 0.02	94.95 ± 0.79	0.44 ± 0.02	0.06 ± 0.02	98.77 ± 0.34	0.63 ± 0.05	0.08 ± 0.04
	SPACR_EFF	90.38 ± 1.47	0.40 ± 0.02	0.19 ± 0.03	95.42 ± 0.32	0.56 ± 0.03	0.26 ± 0.05	98.95 ± 0.46	1.25 ± 0.17	0.58 ± 0.08
	SPACR	90.48 ± 0.85	0.35 ± 0.01	0.04 ± 0.01	95.31 ± 0.80	0.43 ± 0.01	0.05 ± 0.01	99.02 ± 0.33	0.65 ± 0.05	0.07 ± 0.01
Drift	SPACR_VAL	89.81 ± 0.82	30.70 ± 8.50	1.85 ± 2.09	94.77 ± 0.53	40.32 ± 11.34	2.44 ± 2.76	99.11 ± 0.20	71.69 ± 18.68	4.43 ± 5.08
	SPACR_EFF	90.14 ± 0.72	96.38 ± 22.39	376.23 ± 74.19	94.96 ± 0.80	144.93 ± 28.60	566.86 ± 96.00	99.16 ± 0.25	274.57 ± 33.71	1080.45 ± 152.77
	SPACR	90.03 ± 1.29	22.51 ± 2.28	13.04 ± 3.21	95.17 ± 0.71	27.59 ± 2.98	16.02 ± 4.19	99.27 ± 0.28	41.19 ± 5.90	24.02 ± 7.20
UTK Face	SPACR_VAL	90.20 ± 0.62	21.63 ± 3.95	24.15 ± 15.76	94.97 ± 0.29	27.87 ± 5.06	31.32 ± 20.44	98.98 ± 0.11	44.15 ± 6.77	49.69 ± 31.53
	SPACR_EFF	89.81 ± 0.73	127.94 ± 54.67	180.93 ± 79.69	94.63 ± 0.44	210.41 ± 101.23	296.36 ± 144.85	98.84 ± 0.27	391.71 ± 196.21	551.08 ± 279.93
	SPACR	90.13 ± 0.46	21.84 ± 0.84	12.18 ± 1.61	95.02 ± 0.30	27.49 ± 0.98	15.36 ± 2.26	98.95 ± 0.14	42.51 ± 2.06	23.82 ± 4.13

Marginal Coverage: As expected, all variants successfully achieve the target marginal coverage across the different confidence levels. This consistency is due to the post-hoc ICP calibration step, which enforces distribution-free validity regardless of the underlying training objective.

Impact of the Validity Term: The empirical results demonstrate the necessity of the validity penalty. When $\mathcal{L}_{Validity}$ is removed (SPACR_EFF), the model struggles to produce tight and stable prediction intervals. This degradation is most prominent in image datasets such as Drift and UTK Face, where SPACR_EFF exhibits drastically higher median interval widths and IQR compared to the full model. This confirms that dynamically penalizing empirical coverage violations during training is critical for forcing the model to produce efficient interval predictions prior to the post-hoc calibration.

Impact of the Efficiency Term: A close examination of SPACR_VAL reveals the critical role of the efficiency penalty in maintaining conditional adaptivity. While SPACR_VAL yields valid marginal coverage, it frequently results in an IQR of exactly zero (e.g., California and Medical Charges) or near-zero values. This indicates a model collapse where it predicts a rigid, constant interval width for all instances, effectively degrading into a non-adaptive conformal regressor. The inclusion of $\mathcal{L}_{Efficiency}$ in the full SPACR model provides the necessary optimization pressure to shrink intervals for predictable samples, restoring a healthy variance in interval widths that properly scales with instance difficulty.

These results confirm that by jointly optimizing for accuracy, efficiency, and validity, SPACR consistently delivers tight median interval widths while successfully adapting to local heteroscedasticity, ensuring that uncertainty estimates are both valid and highly instance-specific.

5.5 Computation Time Analysis

Experiments were conducted on Ubuntu 24.04.3 LTS with an Intel Core i9-14900KF CPU, an NVIDIA GeForce RTX 5090 GPU, and 32GB of RAM. The timing results are shown in Table 4.

Table 4: Computation training time (in seconds) across datasets. "–" indicates no retraining needed.

Dataset	Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	Total (s)	Dataset	Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	Total (s)
	SICP	28.80 ± 0.34	–	–	28.80 ± 0.34		SICP	19.93 ± 0.24	–	–	19.93 ± 0.24
	NICP	31.29 ± 0.62	–	–	31.29 ± 0.62		NICP	21.33 ± 0.40	–	–	21.33 ± 0.40
Bike Sharing	CQR	34.31 ± 0.39	34.27 ± 0.45	33.94 ± 0.70	102.51 ± 1.05	**Brazilian Houses**	CQR	22.86 ± 0.20	22.84 ± 0.68	22.58 ± 0.47	68.29 ± 0.95
	DOICR	35.42 ± 0.56	35.10 ± 0.66	35.19 ± 0.86	105.71 ± 1.35		DOICR	23.42 ± 0.60	23.76 ± 0.22	23.78 ± 0.48	70.96 ± 1.06
	SPACR	32.80 ± 0.83	–	–	32.80 ± 0.83		SPACR	22.12 ± 0.47	–	–	22.12 ± 0.47
	SICP	33.96 ± 0.75	–	–	33.96 ± 0.75		SICP	16.38 ± 0.44	–	–	16.38 ± 0.44
	NICP	36.76 ± 0.39	–	–	36.76 ± 0.39		NICP	17.17 ± 0.31	–	–	17.17 ± 0.31
California	CQR	39.89 ± 0.95	39.94 ± 0.34	39.62 ± 0.65	119.45 ± 0.85	**CPU Act**	CQR	18.39 ± 0.44	18.65 ± 0.16	18.50 ± 0.35	55.54 ± 0.59
	DOICR	40.81 ± 0.39	41.52 ± 0.88	40.56 ± 0.23	122.89 ± 1.15		DOICR	19.24 ± 0.48	19.07 ± 0.29	19.20 ± 0.31	57.52 ± 0.33
	SPACR	37.99 ± 0.62	–	–	37.99 ± 0.62		SPACR	17.38 ± 0.31	–	–	17.38 ± 0.31
	SICP	83.49 ± 2.05	–	–	83.49 ± 2.05		SICP	30.08 ± 0.44	–	–	30.08 ± 0.44
	NICP	91.84 ± 2.58	–	–	91.84 ± 2.58		NICP	33.12 ± 0.65	–	–	33.12 ± 0.65
Diamonds	CQR	97.91 ± 1.37	99.97 ± 3.44	99.12 ± 3.17	297.00 ± 7.00	**Fifa**	CQR	35.28 ± 0.62	36.05 ± 1.59	35.24 ± 0.45	106.57 ± 1.41
	DOICR	104.67 ± 3.23	103.97 ± 4.24	105.06 ± 4.45	313.70 ± 11.53		DOICR	38.11 ± 3.86	37.19 ± 0.78	37.15 ± 0.43	112.46 ± 4.19
	SPACR	95.11 ± 2.85	–	–	95.11 ± 2.85		SPACR	33.55 ± 0.83	–	–	33.55 ± 0.83
	SICP	36.04 ± 0.51	–	–	36.04 ± 0.51		SICP	22.42 ± 1.09	–	–	22.42 ± 1.09
	NICP	39.00 ± 1.18	–	–	39.00 ± 1.18		NICP	23.91 ± 1.09	–	–	23.91 ± 1.09
House Sales	CQR	41.81 ± 1.04	42.34 ± 1.05	41.94 ± 0.83	126.09 ± 1.70	**Isolet**	CQR	26.62 ± 0.59	26.49 ± 0.79	26.60 ± 0.85	79.70 ± 1.54
	DOICR	43.69 ± 0.97	43.72 ± 0.69	43.67 ± 0.25	131.08 ± 1.46		DOICR	27.49 ± 1.02	28.38 ± 0.94	28.05 ± 1.46	83.92 ± 0.35
	SPACR	40.38 ± 0.95	–	–	40.38 ± 0.95		SPACR	25.86 ± 1.38	–	–	25.86 ± 1.38
	SICP	494.57 ± 24.55	–	–	494.57 ± 24.55		SICP	30.07 ± 4.50	–	–	30.07 ± 4.50
	NICP	567.78 ± 46.81	–	–	567.78 ± 46.81		NICP	31.81 ± 2.79	–	–	31.81 ± 2.79
Medical Charges	CQR	573.16 ± 34.55	581.53 ± 33.07	573.08 ± 38.54	1727.76 ± 89.51	**Pol**	CQR	34.14 ± 2.44	32.26 ± 0.77	32.84 ± 0.80	99.24 ± 3.12
	DOICR	564.86 ± 41.04	554.62 ± 16.41	555.47 ± 29.66	1674.96 ± 45.57		DOICR	34.18 ± 0.29	33.44 ± 0.98	33.93 ± 1.47	101.55 ± 2.16
	SPACR	582.27 ± 56.20	–	–	582.27 ± 56.20		SPACR	31.11 ± 0.96	–	–	31.11 ± 0.96
	SICP	35.00 ± 0.55	–	–	35.00 ± 0.55		SICP	13.55 ± 0.28	–	–	13.55 ± 0.28
	NICP	38.61 ± 0.79	–	–	38.61 ± 0.79		NICP	14.44 ± 0.28	–	–	14.44 ± 0.28
Superconduct	CQR	41.72 ± 0.28	41.55 ± 0.63	40.79 ± 0.66	124.06 ± 0.80	**Wine Quality**	CQR	15.47 ± 0.56	15.49 ± 0.32	15.74 ± 1.12	46.70 ± 1.57
	DOICR	43.23 ± 0.69	42.89 ± 0.73	42.18 ± 0.51	128.30 ± 1.36		DOICR	16.41 ± 0.20	16.14 ± 0.33	16.06 ± 0.81	48.62 ± 0.53
	SPACR	38.99 ± 0.82	–	–	38.99 ± 0.82		SPACR	14.97 ± 0.23	–	–	14.97 ± 0.23
	SICP	1922.86 ± 59.17	–	–	1922.86 ± 59.17		SICP	2940.27 ± 109.72	–	–	2940.27 ± 109.72
	NICP	1918.58 ± 65.54	–	–	1918.58 ± 65.54		NICP	2934.61 ± 34.01	–	–	2934.61 ± 34.01
Drift	CQR	1913.16 ± 52.41	1910.51 ± 77.94	1919.30 ± 115.10	5742.97 ± 178.93	**UTK Face**	CQR	2924.74 ± 170.04	2995.77 ± 40.08	2934.18 ± 123.56	8854.69 ± 223.88
	DOICR	1996.19 ± 181.77	2108.30 ± 334.54	2016.14 ± 210.32	6120.63 ± 643.41		DOICR	3022.59 ± 104.95	3015.58 ± 86.37	2964.23 ± 132.39	9002.39 ± 263.98
	SPACR	1928.90 ± 69.87	–	–	1928.90 ± 69.87		SPACR	3024.50 ± 93.79	–	–	3024.50 ± 93.79

Timing results highlight SPACR's single-pass efficiency. While post-hoc methods (SICP, NICP) lack conformal training, approaches like CQR and DOICR scale linearly ($\mathcal{O}(k)$) by requiring retraining for each of the k confidence levels α. In contrast, SPACR inherently supports multiple α values in constant time ($\mathcal{O}(1)$). For example, on the UTK Face dataset, SPACR requires roughly 50 minutes to evaluate three α values, whereas CQR and DOICR take nearly 2.5 hours, which is a $\sim 3\times$ computational cost. This efficiency gap widens in high-dimensional tasks like bounding box regression in $\mathbb{R}^4$ Mukama et al. (2024), making SPACR highly suited for production environments needing dynamic calibration without prohibitive retraining costs.

6 Conclusion

In this paper, we introduced SPACR (Single-Pass Adaptive Conformal Regressor), a novel framework for directly training uncertainty-aware regressors via Conformal Prediction. By integrating accuracy, efficiency, and validity into a single differentiable loss, SPACR eliminates batch-splitting and fixed training confidence levels, which gives it a computational advantage over other methods. Experiments show that a single SPACR model efficiently computes intervals across arbitrary confidence levels, outperforming standard ICP, normalized ICP, and CQR as well as and DOICR, a state-of-the-art method for training conformal regressors, by yielding tighter bounds without sacrificing coverage.

While we focused here on a specific form of $\mathcal{L}_{\text{Validity}}$, the fact that it is related to robust losses (Vapnik et al. (1996)) indicates that we could probably take inspiration from other robust forms of loss functions in the literature (Barron (2019); Ghosh et al. (2017)). We could also think of including some asymmetry in those losses, as done for the pinball loss function (Huang et al. (2013)), for which some proposals already exist (Hu et al. (2012)). However, this would probably require a better understanding of the interplay between the different losses used in our proposal, an aspect we intend to tackle in future work.

References

Frédéric Allaire, Vivien Mallet, and Jean-Baptiste Filippi. Novel method for a posteriori uncertainty quantification in wildland fire spread simulation. *Applied Mathematical Modelling*, 90:527–546, 2021.

Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

Jonathan T Barron. A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4331–4339, 2019.

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

Nicolo Colombo and Vladimir Vovk. Training conformal predictors. In *Conformal and Probabilistic Prediction and Applications*, pp. 55–64. PMLR, 2020.

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. The limits of distribution-free conditional predictive inference. *Information and Inference: A Journal of the IMA*, 10(2):455–482, 2021.

Aritra Ghosh, Himanshu Kumar, and P Shanti Sastry. Robust loss functions under label noise for deep neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? *Advances in neural information processing systems*, 35:507–520, 2022.

Shasha Han and Paulin Coulibaly. Bayesian flood forecasting methods: A review. *Journal of Hydrology*, 551:340–351, 2017.

Ting Hu, Dao-Hong Xiang, and Ding-Xuan Zhou. Online learning for quantile regression and support vector regression. *Journal of Statistical Planning and Inference*, 142(12):3107–3122, 2012.

Xiaolin Huang, Lei Shi, and Johan AK Suykens. Support vector machine classifier with pinball loss. *IEEE transactions on pattern analysis and machine intelligence*, 36(5):984–997, 2013.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

Haocheng Lei and Anthony Bellotti. Reliable prediction intervals with directly optimized inductive conformal regression for deep learning. *Neural Networks*, 168:194–205, 2023.

Sizhuo Li, Dimitri Gominski, Martin Brandt, Xiaoye Tong, and Philippe Ciais. Get your embedding space in order: Domain-adaptive regression for forest monitoring. In *European Conference on Computer Vision*, pp. 94–111. Springer, 2024.

Rhiannon Michelmore, Matthew Wicker, Luca Laurenti, Luca Cardelli, Yarin Gal, and Marta Kwiatkowska. Uncertainty quantification with statistical guarantees in end-to-end autonomous driving control. In *2020 IEEE international conference on robotics and automation (ICRA)*, pp. 7344–7350. IEEE, 2020.

Bruce Cyusa Mukama, Soundouss Messoudi, Sylvain Rousseau, and Sébastien Destercke. Copula-based conformal prediction for object detection: a more efficient approach. *Proceedings of Machine Learning Research*, 230:1–18, 2024.

Harris Papadopoulos. *Inductive conformal prediction: Theory and application to neural networks*. INTECH Open Access Publisher Rijeka, 2008.

Harris Papadopoulos and Haris Haralambous. Reliable prediction intervals with regression neural networks. *Neural Networks*, 24(8):842–851, 2011.

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

Silvia Seoni, Vicnesh Jahmunah, Massimo Salvi, Prabal Datta Barua, Filippo Molinari, and U Rajendra Acharya. Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013–2023). *Computers in Biology and Medicine*, 165:107441, 2023.

David Stutz, Krishnamurthy Dvijotham, Ali Taylan Cemgil, and Arnaud Doucet. Learning optimal conformal classifiers. In *International Conference on Learning Representations*, 2022.

Vladimir Vapnik, Steven Golowich, and Alex Smola. Support vector method for function approximation, regression estimation and signal processing. *Advances in neural information processing systems*, 9, 1996.

Janette Vazquez and Julio C Facelli. Conformal prediction in clinical medical sciences. *Journal of Healthcare Informatics Research*, 6(3):241–252, 2022.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer Science & Business Media, 2005.

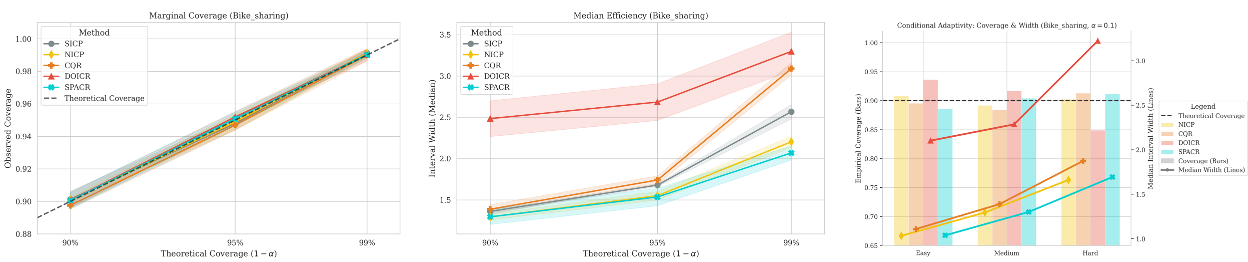
Christopher Yeh, Nicolas Christianson, Adam Wierman, and Yisong Yue. Conformal risk training: End-to-end optimization of conformal risk control. *arXiv preprint arXiv:2510.08748*, 2025.

Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5810–5818, 2017.

A Additional Results

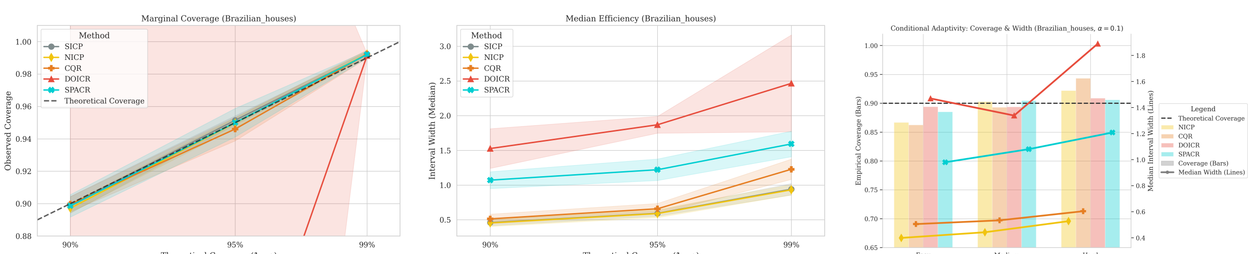
A.1 Performance (coverage and efficiency)

Figures 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18 detail the coverage and efficiency metrics for the remaining datasets. These results reinforce the main findings of the study: while all methods generally achieve the desired coverage, SPACR consistently produces the narrowest and most stable prediction intervals with robust conditional adaptivity.



(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 7: Performance figures for the different approaches for the Bike Sharing dataset.

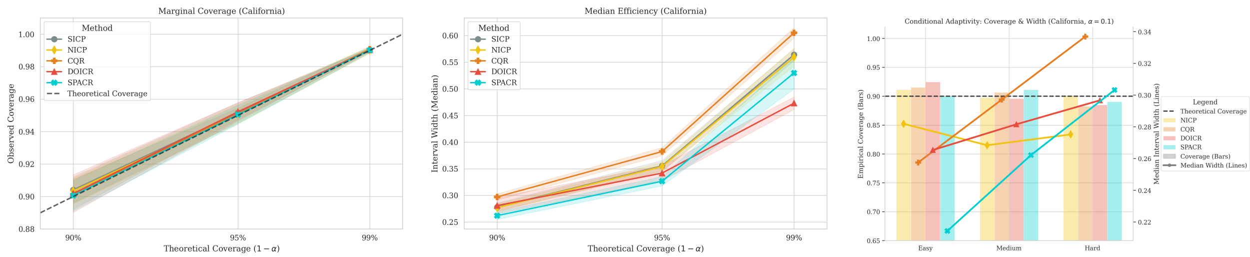


(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 8: Performance figures for the different approaches for the Brazilian Houses dataset.

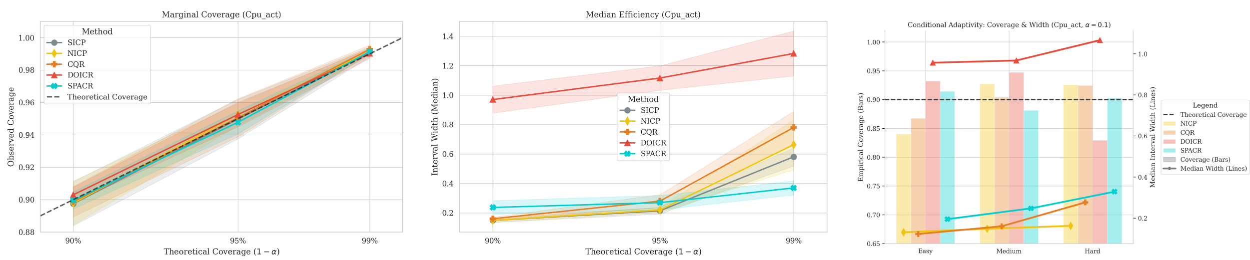
A.2 Sensitivity Analysis of the Regularization Parameter λ

Tables 5 and 6 and Figures 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 and 29 illustrate how SPACR balances coverage and interval efficiency across different values of the regularization parameter λ on the remaining datasets. These visualizations confirm the main findings that intermediate values of λ strike an effective trade-off, producing compact, well-calibrated intervals. These patterns reinforce the tuning dynamics described in the



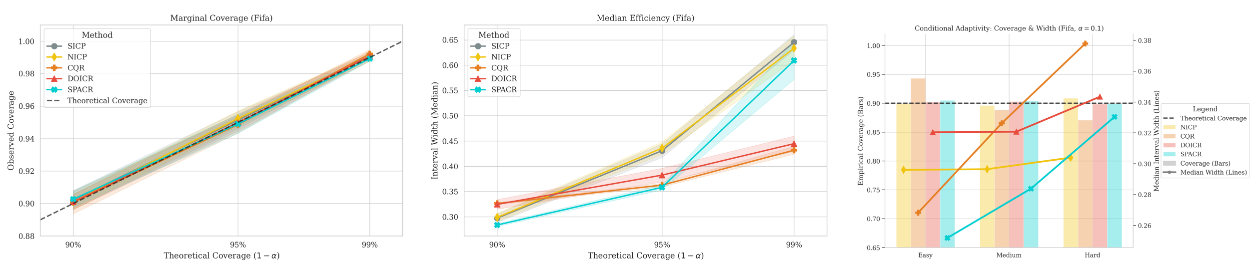
(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 9: Performance figures for the different approaches for the California dataset.



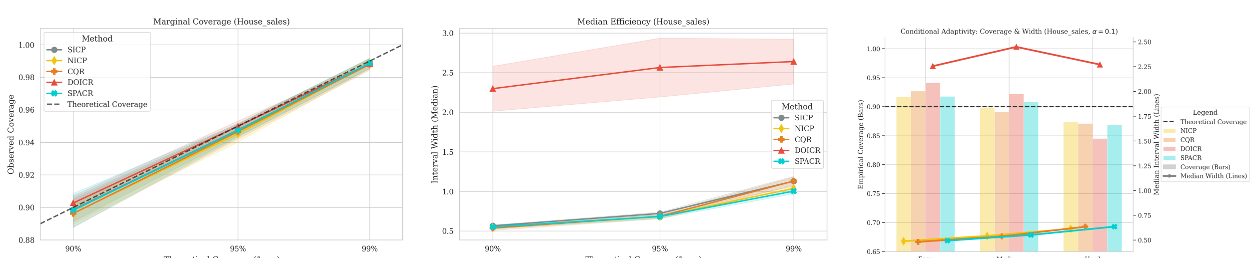
(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 10: Performance figures for the different approaches for the Cpu Act dataset.



(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

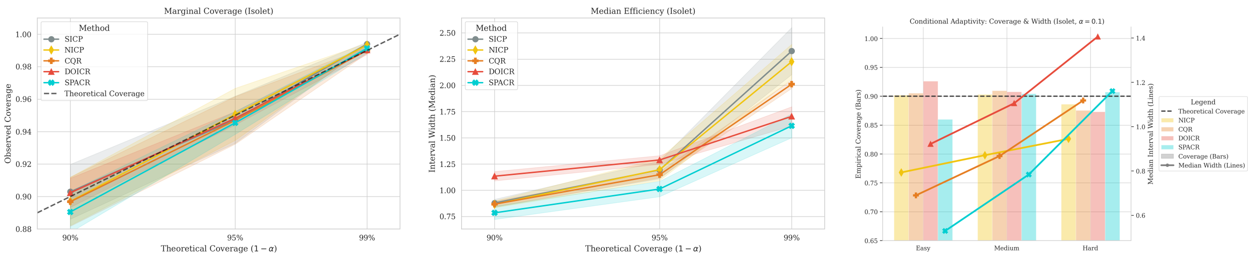
Figure 11: Performance figures for the different approaches for the Fifa dataset.



(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 12: Performance figures for the different approaches for the House Sales dataset.

main paper. As a result, one can simply tune λ using a small holdout validation set to measure coverage and interval width at the desired target α and sweep across $\lambda \in [1, 50]$ values to find the best one.

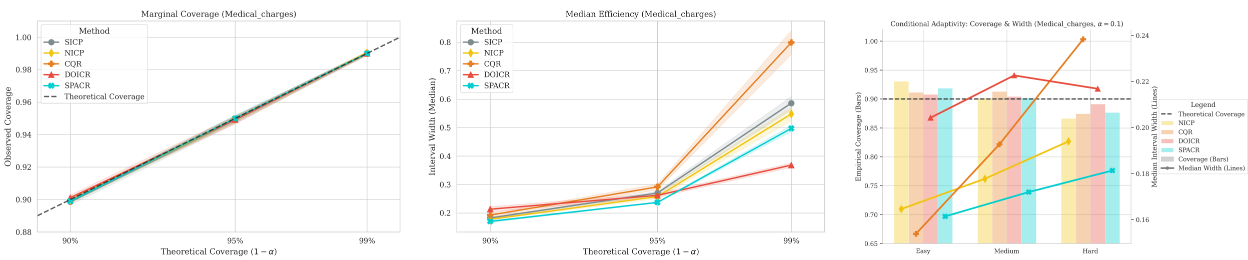


(a) Coverage vs. calibration line.

(b) Median efficiency.

(c) Conditional coverage & efficiency.

Figure 13: Performance figures for the different approaches for the Isolet dataset.

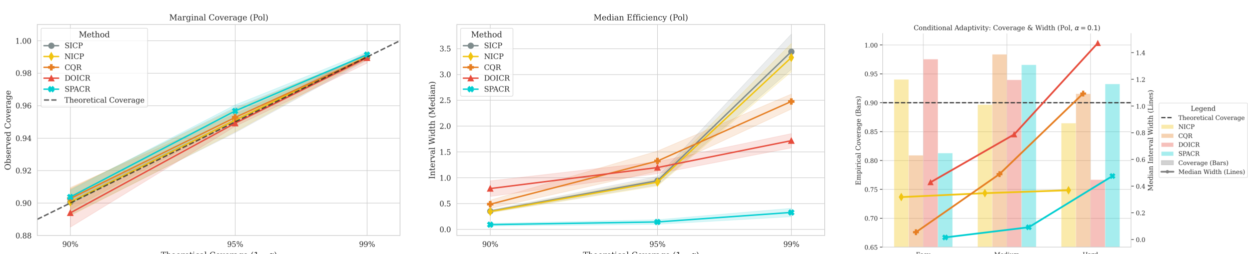


(a) Coverage vs. calibration line.

(b) Median efficiency.

(c) Conditional coverage & efficiency.

Figure 14: Performance figures for the different approaches for the Medical Charges dataset.

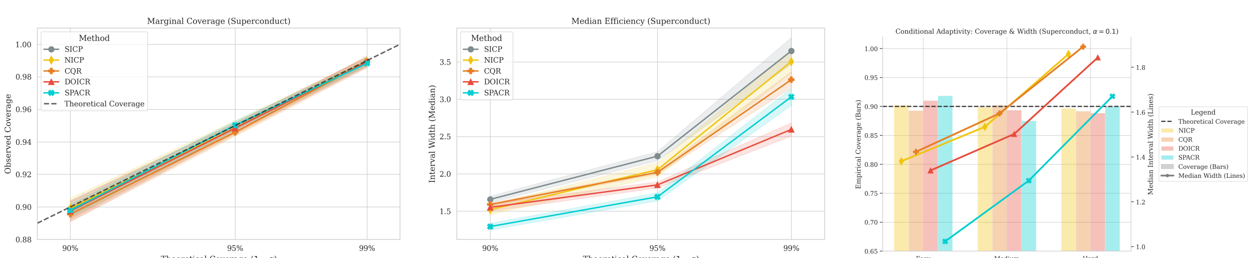


(a) Coverage vs. calibration line.

(b) Median efficiency.

(c) Conditional coverage & efficiency.

Figure 15: Performance figures for the different approaches for the Pol dataset.

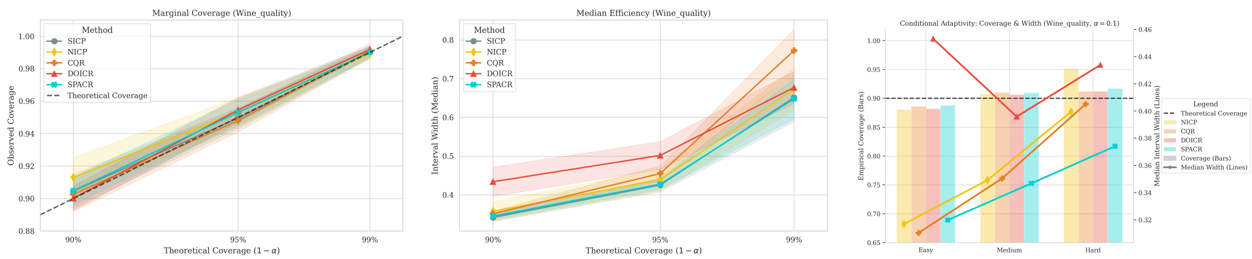


(a) Coverage vs. calibration line.

(b) Median efficiency.

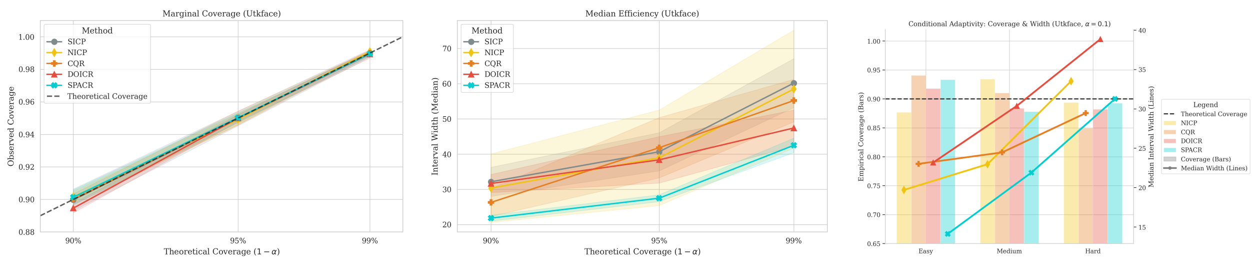
(c) Conditional coverage & efficiency.

Figure 16: Performance figures for the different approaches for the Superconduct dataset.



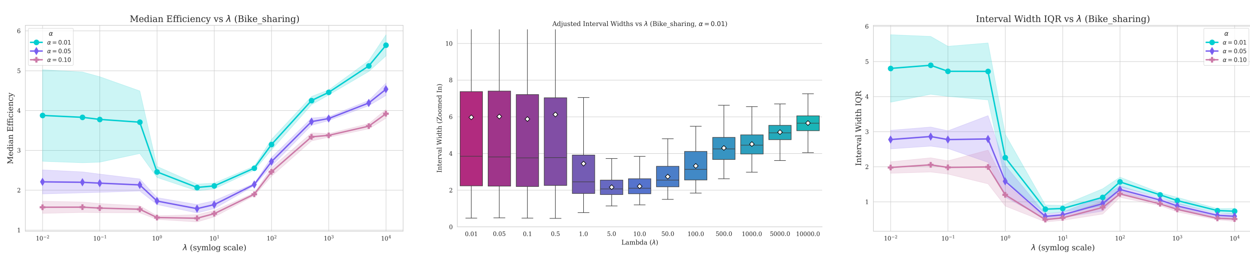
(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 17: Performance figures for the different approaches for the Wine Quality dataset.



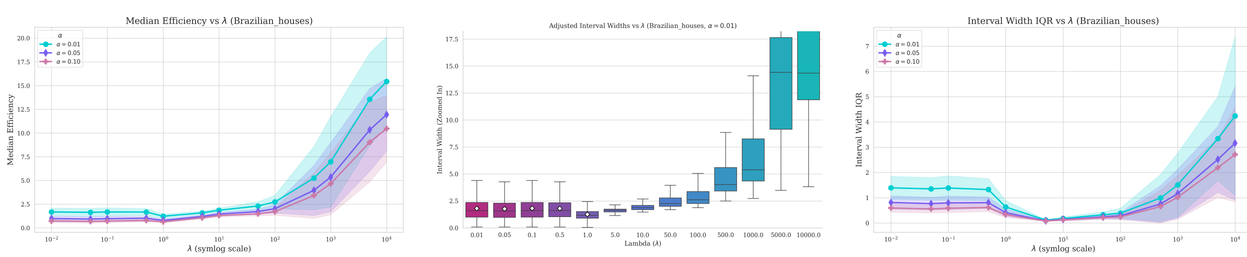
(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 18: Performance figures for the different approaches for the Utkface dataset.



(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 19: Results analysis for varying λ values for the Bike Sharing dataset.



(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

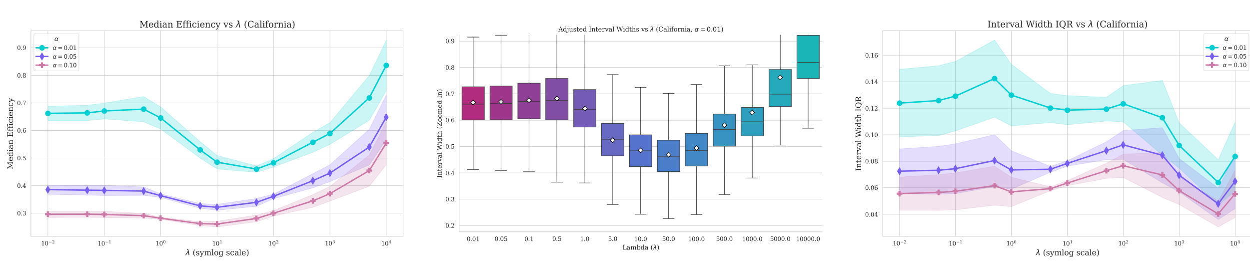
Figure 20: Results analysis for varying λ values for the Brazilian Houses dataset.

Table 5: Impact of λ on Marginal Coverage, Median Efficiency, and Interval Width IQR. Best results per dataset and confidence level are highlighted in bold. The row corresponding to $\lambda = 5$ is shaded in gray.

Dataset	Lambda (λ)	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
		Cov. (%)	Med. Eff.	IQR	Cov. (%)	Med. Eff.	IQR	Cov. (%)	Med. Eff.	IQR
Bike Sharing	0.01	90.10 ± 0.25	1.57 ± 0.15	1.98 ± 0.16	95.51 ± 0.61	2.21 ± 0.30	2.77 ± 0.26	99.04 ± 0.16	3.88 ± 1.15	4.80 ± 0.96
	0.05	90.25 ± 0.18	1.57 ± 0.13	2.05 ± 0.20	95.42 ± 0.41	2.20 ± 0.26	2.86 ± 0.27	99.02 ± 0.18	3.83 ± 1.14	4.89 ± 0.82
	0.1	90.32 ± 0.21	1.55 ± 0.12	1.98 ± 0.19	95.48 ± 0.50	2.18 ± 0.23	2.77 ± 0.25	99.06 ± 0.18	3.78 ± 1.07	4.72 ± 0.71
	0.5	**89.99 ± 0.16**	1.52 ± 0.09	1.99 ± 0.48	95.54 ± 0.42	2.13 ± 0.15	2.79 ± 0.67	99.15 ± 0.22	3.71 ± 0.79	4.72 ± 0.81
	1	89.95 ± 0.71	1.31 ± 0.05	1.20 ± 0.32	95.11 ± 0.44	1.73 ± 0.09	1.59 ± 0.45	99.22 ± 0.13	2.46 ± 0.14	2.26 ± 0.66
	5	90.06 ± 0.52	**1.29 ± 0.09**	**0.49 ± 0.06**	95.09 ± 0.41	**1.54 ± 0.11**	**0.58 ± 0.07**	99.02 ± 0.22	**2.07 ± 0.09**	0.79 ± 0.12
	10	90.12 ± 0.80	1.41 ± 0.08	0.54 ± 0.07	95.07 ± 0.52	1.64 ± 0.10	0.63 ± 0.08	98.99 ± 0.21	2.11 ± 0.07	0.81 ± 0.09
	50	90.29 ± 0.49	1.90 ± 0.03	0.84 ± 0.19	95.06 ± 0.59	2.14 ± 0.03	0.95 ± 0.21	99.02 ± 0.25	2.55 ± 0.05	1.13 ± 0.25
	100	90.25 ± 0.72	2.46 ± 0.11	1.22 ± 0.10	94.90 ± 0.25	2.72 ± 0.11	1.35 ± 0.11	98.98 ± 0.42	3.15 ± 0.13	1.57 ± 0.14
	500	89.99 ± 0.40	3.34 ± 0.10	0.94 ± 0.05	94.97 ± 0.50	3.72 ± 0.09	1.05 ± 0.05	**99.00 ± 0.24**	4.25 ± 0.10	1.20 ± 0.06
	1000	89.94 ± 0.53	3.38 ± 0.05	0.78 ± 0.07	**94.97 ± 0.34**	3.80 ± 0.05	0.88 ± 0.08	99.03 ± 0.37	4.46 ± 0.06	1.04 ± 0.10
	5000	89.59 ± 0.77	3.61 ± 0.07	0.53 ± 0.04	95.05 ± 0.51	4.19 ± 0.07	0.61 ± 0.05	99.11 ± 0.26	5.12 ± 0.14	**0.75 ± 0.06**
	10000	89.76 ± 0.46	3.92 ± 0.11	0.51 ± 0.07	94.80 ± 0.51	4.54 ± 0.16	0.59 ± 0.07	99.10 ± 0.16	5.64 ± 0.27	0.73 ± 0.09
Brazilian Houses	0.01	90.12 ± 1.25	0.73 ± 0.17	0.59 ± 0.17	95.12 ± 0.54	0.99 ± 0.27	0.81 ± 0.27	99.10 ± 0.23	1.69 ± 0.42	1.39 ± 0.46
	0.05	**90.01 ± 0.84**	0.69 ± 0.17	0.56 ± 0.16	95.21 ± 0.54	0.94 ± 0.26	0.77 ± 0.25	99.16 ± 0.22	1.65 ± 0.43	1.35 ± 0.45
	0.1	90.11 ± 0.83	0.72 ± 0.16	0.58 ± 0.17	95.22 ± 0.41	0.98 ± 0.27	0.80 ± 0.27	99.15 ± 0.23	1.69 ± 0.44	1.39 ± 0.48
	0.5	90.20 ± 0.83	0.79 ± 0.14	0.61 ± 0.16	95.07 ± 0.42	1.03 ± 0.21	0.81 ± 0.23	99.14 ± 0.32	1.68 ± 0.37	1.32 ± 0.44
	1	90.12 ± 0.16	**0.65 ± 0.11**	0.33 ± 0.12	94.86 ± 0.37	**0.80 ± 0.14**	0.41 ± 0.15	99.06 ± 0.32	**1.23 ± 0.23**	0.64 ± 0.26
	5	89.86 ± 0.69	1.07 ± 0.12	**0.08 ± 0.01**	**95.00 ± 0.88**	1.22 ± 0.15	**0.09 ± 0.01**	99.21 ± 0.22	1.59 ± 0.18	**0.12 ± 0.02**
	10	90.14 ± 0.52	1.30 ± 0.13	0.12 ± 0.04	95.29 ± 0.74	1.47 ± 0.16	0.14 ± 0.05	99.22 ± 0.18	1.87 ± 0.22	0.18 ± 0.06
	50	90.13 ± 0.89	1.49 ± 0.21	0.21 ± 0.07	94.98 ± 0.50	1.73 ± 0.25	0.24 ± 0.08	**99.01 ± 0.44**	2.31 ± 0.46	0.33 ± 0.12
	100	90.07 ± 0.56	1.72 ± 0.38	0.24 ± 0.11	95.18 ± 0.39	2.03 ± 0.45	0.29 ± 0.14	98.96 ± 0.40	2.74 ± 0.70	0.39 ± 0.20
	500	90.33 ± 0.99	3.41 ± 2.41	0.65 ± 0.65	95.06 ± 0.90	3.96 ± 2.64	0.75 ± 0.73	99.13 ± 0.34	5.25 ± 3.34	0.99 ± 0.93
	1000	90.43 ± 0.80	4.67 ± 3.28	1.02 ± 0.87	95.27 ± 0.54	5.36 ± 3.64	1.17 ± 0.97	99.12 ± 0.31	6.97 ± 4.76	1.51 ± 1.27
	5000	90.48 ± 0.81	9.04 ± 4.18	2.19 ± 1.18	95.18 ± 0.67	10.34 ± 4.36	2.52 ± 1.30	99.20 ± 0.20	13.57 ± 4.89	3.34 ± 1.67
	10000	90.49 ± 0.97	10.46 ± 3.53	2.71 ± 1.87	95.13 ± 0.75	11.95 ± 3.84	3.16 ± 2.25	99.15 ± 0.04	15.43 ± 4.74	4.24 ± 3.15
California	0.01	90.46 ± 1.09	0.30 ± 0.01	0.06 ± 0.01	95.38 ± 0.75	0.39 ± 0.02	0.07 ± 0.02	99.09 ± 0.18	0.66 ± 0.03	0.12 ± 0.03
	0.05	90.60 ± 0.92	0.30 ± 0.01	0.06 ± 0.01	95.30 ± 0.77	0.38 ± 0.02	0.07 ± 0.02	99.09 ± 0.20	0.66 ± 0.03	0.13 ± 0.03
	0.1	90.63 ± 0.84	0.30 ± 0.01	0.06 ± 0.01	95.29 ± 0.75	0.38 ± 0.02	0.07 ± 0.02	99.10 ± 0.15	0.67 ± 0.03	0.13 ± 0.03
	0.5	90.53 ± 0.89	0.29 ± 0.01	0.06 ± 0.01	95.43 ± 0.69	0.38 ± 0.02	0.08 ± 0.02	99.13 ± 0.05	0.68 ± 0.05	0.14 ± 0.03
	1	90.40 ± 0.96	0.28 ± 0.00	0.06 ± 0.01	95.20 ± 0.48	0.36 ± 0.01	0.07 ± 0.01	99.07 ± 0.05	0.65 ± 0.04	0.13 ± 0.02
	5	**90.06 ± 0.94**	**0.26 ± 0.01**	0.06 ± 0.00	95.04 ± 0.57	0.33 ± 0.01	0.07 ± 0.00	**99.02 ± 0.15**	0.53 ± 0.03	0.12 ± 0.01
	10	89.80 ± 1.13	**0.26 ± 0.01**	0.06 ± 0.00	**95.00 ± 0.66**	**0.32 ± 0.01**	0.08 ± 0.00	98.95 ± 0.15	0.48 ± 0.02	0.12 ± 0.01
	50	90.07 ± 0.70	0.28 ± 0.01	0.07 ± 0.01	94.86 ± 0.53	0.34 ± 0.01	0.09 ± 0.01	99.02 ± 0.15	**0.46 ± 0.01**	0.12 ± 0.01
	100	90.36 ± 0.42	0.30 ± 0.01	0.08 ± 0.01	95.11 ± 0.50	0.36 ± 0.01	0.09 ± 0.01	99.06 ± 0.15	0.48 ± 0.02	0.12 ± 0.01
	500	90.65 ± 0.51	0.34 ± 0.02	0.07 ± 0.02	95.49 ± 0.52	0.42 ± 0.03	0.08 ± 0.02	99.08 ± 0.17	0.56 ± 0.04	0.11 ± 0.03
	1000	90.65 ± 0.31	0.37 ± 0.03	0.06 ± 0.01	95.38 ± 0.51	0.45 ± 0.03	0.07 ± 0.01	99.09 ± 0.06	0.59 ± 0.04	0.09 ± 0.02
	5000	90.63 ± 0.46	0.45 ± 0.06	**0.04 ± 0.01**	95.52 ± 0.33	0.54 ± 0.06	**0.05 ± 0.01**	99.10 ± 0.09	0.72 ± 0.08	**0.06 ± 0.02**
	10000	90.74 ± 0.46	0.55 ± 0.08	0.06 ± 0.02	95.45 ± 0.25	0.65 ± 0.08	0.06 ± 0.02	99.14 ± 0.18	0.84 ± 0.09	0.08 ± 0.03
Diamonds	0.01	89.88 ± 0.41	0.54 ± 0.03	0.62 ± 0.08	94.98 ± 0.29	0.84 ± 0.05	0.97 ± 0.14	99.07 ± 0.09	1.75 ± 0.06	2.01 ± 0.21
	0.05	89.84 ± 0.43	0.53 ± 0.03	0.59 ± 0.08	94.93 ± 0.35	0.82 ± 0.05	0.92 ± 0.14	99.09 ± 0.08	1.72 ± 0.07	1.93 ± 0.22
	0.1	89.86 ± 0.39	0.53 ± 0.03	0.59 ± 0.09	**95.00 ± 0.29**	0.81 ± 0.06	0.92 ± 0.15	99.06 ± 0.09	1.72 ± 0.08	1.93 ± 0.25
	0.5	89.84 ± 0.34	0.53 ± 0.02	0.60 ± 0.04	95.03 ± 0.33	0.82 ± 0.03	0.94 ± 0.07	99.03 ± 0.12	1.74 ± 0.15	2.00 ± 0.24
	1	89.85 ± 0.30	0.45 ± 0.02	0.38 ± 0.04	95.10 ± 0.31	0.67 ± 0.04	0.57 ± 0.06	99.03 ± 0.09	1.35 ± 0.10	1.14 ± 0.14
	5	90.05 ± 0.36	**0.34 ± 0.00**	0.06 ± 0.00	95.02 ± 0.11	**0.42 ± 0.00**	0.07 ± 0.01	99.00 ± 0.08	**0.60 ± 0.01**	0.10 ± 0.01
	10	89.90 ± 0.51	0.36 ± 0.00	**0.05 ± 0.00**	95.04 ± 0.19	0.42 ± 0.01	**0.05 ± 0.00**	98.98 ± 0.13	**0.60 ± 0.01**	**0.08 ± 0.01**
	50	89.81 ± 0.27	0.40 ± 0.03	0.05 ± 0.01	94.76 ± 0.19	0.47 ± 0.04	0.06 ± 0.01	98.97 ± 0.10	0.61 ± 0.04	**0.08 ± 0.01**
	100	89.86 ± 0.35	0.43 ± 0.05	0.06 ± 0.02	94.76 ± 0.16	0.51 ± 0.06	0.07 ± 0.02	98.97 ± 0.10	0.65 ± 0.07	0.09 ± 0.03
	500	89.85 ± 0.18	0.79 ± 0.34	0.16 ± 0.09	94.99 ± 0.18	0.90 ± 0.35	0.18 ± 0.10	99.02 ± 0.13	1.10 ± 0.39	0.22 ± 0.11
	1000	**90.03 ± 0.27**	0.93 ± 0.35	0.19 ± 0.11	95.01 ± 0.13	1.06 ± 0.38	0.22 ± 0.12	99.05 ± 0.12	1.30 ± 0.44	0.27 ± 0.15
	5000	90.04 ± 0.18	1.34 ± 0.28	0.34 ± 0.18	95.07 ± 0.19	1.51 ± 0.31	0.39 ± 0.20	99.01 ± 0.09	1.87 ± 0.39	0.49 ± 0.26
	10000	90.08 ± 0.41	1.51 ± 0.27	0.40 ± 0.11	95.08 ± 0.36	1.73 ± 0.24	0.47 ± 0.14	**99.00 ± 0.07**	2.22 ± 0.31	0.61 ± 0.21
Fifa	0.01	90.17 ± 0.57	0.34 ± 0.02	0.06 ± 0.02	95.29 ± 0.25	0.49 ± 0.02	0.08 ± 0.03	99.19 ± 0.07	0.96 ± 0.22	0.16 ± 0.07
	0.05	90.17 ± 0.60	0.34 ± 0.02	0.06 ± 0.02	95.31 ± 0.26	0.50 ± 0.02	0.09 ± 0.03	99.16 ± 0.09	0.99 ± 0.24	0.17 ± 0.08
	0.1	90.18 ± 0.64	0.34 ± 0.02	0.06 ± 0.02	95.30 ± 0.26	0.50 ± 0.02	0.09 ± 0.03	99.15 ± 0.09	0.98 ± 0.16	0.18 ± 0.07
	0.5	90.33 ± 0.56	0.36 ± 0.02	0.08 ± 0.02	95.24 ± 0.26	0.53 ± 0.03	0.12 ± 0.03	99.14 ± 0.07	1.22 ± 0.26	0.27 ± 0.10
	1	90.25 ± 0.61	0.35 ± 0.02	0.07 ± 0.02	95.17 ± 0.36	0.52 ± 0.03	0.11 ± 0.03	99.14 ± 0.16	1.13 ± 0.22	0.23 ± 0.09
	5	90.27 ± 0.51	**0.28 ± 0.00**	**0.05 ± 0.00**	94.90 ± 0.57	0.36 ± 0.01	0.07 ± 0.01	98.93 ± 0.16	0.61 ± 0.04	0.12 ± 0.02
	10	**90.01 ± 0.81**	0.31 ± 0.00	0.06 ± 0.00	95.05 ± 0.25	**0.36 ± 0.00**	0.07 ± 0.01	**99.00 ± 0.21**	0.51 ± 0.02	0.10 ± 0.01
	50	89.95 ± 0.59	0.34 ± 0.00	**0.05 ± 0.00**	94.94 ± 0.54	0.38 ± 0.00	**0.06 ± 0.00**	99.01 ± 0.28	**0.44 ± 0.01**	**0.07 ± 0.00**
	100	90.15 ± 0.55	0.35 ± 0.01	**0.05 ± 0.00**	**95.05 ± 0.23**	0.39 ± 0.02	**0.06 ± 0.00**	98.92 ± 0.19	0.46 ± 0.03	**0.07 ± 0.00**
	500	90.09 ± 0.35	0.46 ± 0.06	0.07 ± 0.01	94.87 ± 0.25	0.52 ± 0.07	0.08 ± 0.01	98.88 ± 0.18	0.61 ± 0.07	0.09 ± 0.01
	1000	89.90 ± 0.64	0.54 ± 0.10	0.09 ± 0.02	95.07 ± 0.39	0.60 ± 0.10	0.10 ± 0.02	99.06 ± 0.27	0.70 ± 0.11	0.12 ± 0.03
	5000	89.79 ± 0.78	0.96 ± 0.10	0.14 ± 0.02	95.11 ± 0.48	1.06 ± 0.11	0.16 ± 0.02	99.07 ± 0.37	1.23 ± 0.15	0.18 ± 0.03
	10000	89.85 ± 0.81	1.21 ± 0.19	0.19 ± 0.01	94.91 ± 0.72	1.35 ± 0.21	0.21 ± 0.02	99.04 ± 0.47	1.59 ± 0.24	0.25 ± 0.02
House Sales	0.01	89.77 ± 0.77	0.60 ± 0.02	0.20 ± 0.01	94.79 ± 0.41	0.79 ± 0.03	0.26 ± 0.01	98.91 ± 0.25	1.27 ± 0.05	0.42 ± 0.02
	0.05	89.74 ± 0.62	0.60 ± 0.02	0.20 ± 0.01	94.74 ± 0.43	0.78 ± 0.03	0.26 ± 0.02	98.90 ± 0.30	1.27 ± 0.06	0.42 ± 0.02
	0.1	89.74 ± 0.53	0.60 ± 0.02	0.20 ± 0.01	94.71 ± 0.39	0.78 ± 0.03	0.26 ± 0.01	98.90 ± 0.25	1.27 ± 0.06	0.42 ± 0.02
	0.5	89.82 ± 0.71	0.60 ± 0.02	0.19 ± 0.01	94.83 ± 0.47	0.78 ± 0.02	0.25 ± 0.01	98.94 ± 0.18	1.26 ± 0.04	0.40 ± 0.01
	1	89.84 ± 0.84	0.58 ± 0.02	0.13 ± 0.00	94.74 ± 0.33	0.75 ± 0.02	0.17 ± 0.01	98.86 ± 0.19	1.17 ± 0.06	0.26 ± 0.01
	5	89.83 ± 1.04	**0.55 ± 0.02**	0.10 ± 0.01	94.74 ± 0.41	**0.68 ± 0.03**	0.12 ± 0.01	98.87 ± 0.31	**1.00 ± 0.03**	0.18 ± 0.01
	10	89.78 ± 0.79	0.63 ± 0.05	**0.09 ± 0.01**	94.94 ± 0.33	0.76 ± 0.05	**0.11 ± 0.01**	98.86 ± 0.15	1.05 ± 0.05	**0.16 ± 0.02**
	50	89.59 ± 0.46	1.43 ± 0.17	0.15 ± 0.03	94.99 ± 0.37	1.59 ± 0.18	0.17 ± 0.03	99.07 ± 0.14	1.93 ± 0.22	0.20 ± 0.05
	100	89.56 ± 0.89	1.84 ± 0.32	0.19 ± 0.04	94.80 ± 0.48	2.04 ± 0.32	0.21 ± 0.05	99.10 ± 0.15	2.47 ± 0.33	0.26 ± 0.06
	500	90.11 ± 0.34	2.02 ± 0.29	0.23 ± 0.07	95.13 ± 0.41	2.37 ± 0.31	0.27 ± 0.09	**99.04 ± 0.22**	3.21 ± 0.47	0.37 ± 0.13
	1000	90.57 ± 0.45	2.15 ± 0.26	0.28 ± 0.10	95.34 ± 0.21	2.62 ± 0.35	0.35 ± 0.13	99.07 ± 0.14	3.64 ± 0.52	0.48 ± 0.19
	5000	**90.03 ± 0.62**	3.05 ± 0.47	0.59 ± 0.22	95.32 ± 0.30	3.70 ± 0.51	0.72 ± 0.25	99.05 ± 0.20	4.90 ± 0.69	0.95 ± 0.34
	10000	90.09 ± 0.88	3.70 ± 0.74	0.81 ± 0.30	**94.99 ± 0.45**	4.44 ± 0.90	0.97 ± 0.36	98.95 ± 0.18	5.81 ± 1.12	1.27 ± 0.47

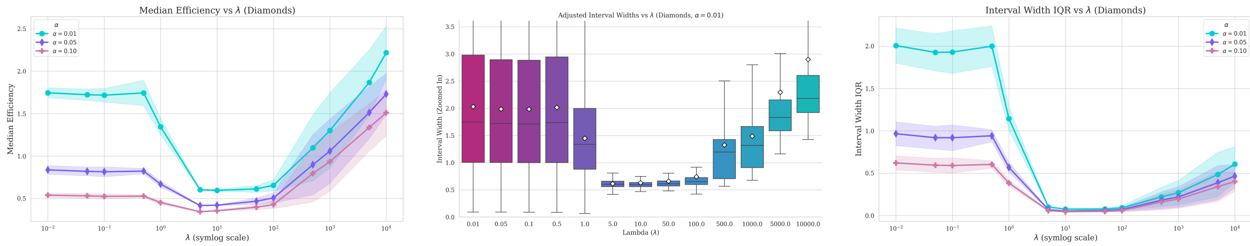
Table 6: Impact of λ on Marginal Coverage, Median Efficiency, and Interval Width IQR. Best results per dataset and confidence level are highlighted in bold. The row corresponding to $\lambda = 5$ is shaded in gray.

Dataset	**Lambda (λ)**	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
		Cov. (%)	Med. Eff.	IQR	Cov. (%)	Med. Eff.	IQR	Cov. (%)	Med. Eff.	IQR
	0.01	90.17 ± 1.13	1.37 ± 0.26	1.84 ± 0.52	95.27 ± 0.62	2.05 ± 0.30	2.76 ± 0.67	99.12 ± 0.22	4.09 ± 0.67	5.50 ± 1.42
	0.05	90.12 ± 1.40	1.35 ± 0.26	1.80 ± 0.50	95.17 ± 0.46	2.00 ± 0.32	2.67 ± 0.66	99.06 ± 0.21	3.94 ± 0.72	5.25 ± 1.42
	0.1	90.23 ± 1.12	1.33 ± 0.23	1.75 ± 0.47	95.36 ± 0.65	2.04 ± 0.37	2.67 ± 0.74	99.14 ± 0.19	3.97 ± 0.67	5.21 ± 1.33
	0.5	90.13 ± 0.68	1.13 ± 0.08	1.17 ± 0.16	95.04 ± 0.81	1.63 ± 0.17	1.69 ± 0.28	99.19 ± 0.17	3.15 ± 0.42	3.26 ± 0.58
	1	90.06 ± 1.02	0.95 ± 0.02	0.74 ± 0.04	94.85 ± 0.87	1.29 ± 0.05	1.00 ± 0.06	99.05 ± 0.22	2.28 ± 0.11	1.77 ± 0.07
	5	89.05 ± 1.26	**0.78 ± 0.06**	0.43 ± 0.03	94.54 ± 0.70	1.01 ± 0.07	0.56 ± 0.05	99.13 ± 0.28	1.61 ± 0.11	0.90 ± 0.09
Isolet	10	89.15 ± 1.10	0.81 ± 0.05	0.44 ± 0.02	93.96 ± 1.34	**1.00 ± 0.09**	0.55 ± 0.03	99.22 ± 0.26	**1.58 ± 0.13**	0.87 ± 0.05
	50	89.83 ± 1.15	1.03 ± 0.05	0.46 ± 0.03	94.91 ± 0.80	1.23 ± 0.07	0.55 ± 0.04	**99.04 ± 0.15**	1.69 ± 0.13	0.75 ± 0.07
	100	90.23 ± 0.67	1.17 ± 0.06	0.45 ± 0.03	95.06 ± 0.79	1.37 ± 0.08	0.53 ± 0.04	99.24 ± 0.08	1.83 ± 0.10	0.71 ± 0.04
	500	**90.03 ± 0.88**	1.45 ± 0.06	**0.37 ± 0.02**	94.68 ± 0.68	1.70 ± 0.09	0.43 ± 0.03	99.18 ± 0.23	2.23 ± 0.13	0.57 ± 0.05
	1000	90.06 ± 0.74	1.59 ± 0.05	0.36 ± 0.04	95.21 ± 0.50	1.86 ± 0.07	**0.42 ± 0.04**	99.14 ± 0.10	2.39 ± 0.12	**0.54 ± 0.05**
	5000	89.26 ± 0.20	2.10 ± 0.09	0.52 ± 0.05	94.94 ± 0.58	2.45 ± 0.07	0.61 ± 0.06	98.94 ± 0.33	3.13 ± 0.11	0.78 ± 0.08
	10000	89.00 ± 0.89	2.43 ± 0.07	0.74 ± 0.09	**95.04 ± 0.59**	2.83 ± 0.04	0.86 ± 0.11	98.82 ± 0.30	3.59 ± 0.10	1.09 ± 0.14
	0.01	90.17 ± 0.68	0.48 ± 0.03	0.63 ± 0.02	95.37 ± 0.71	0.83 ± 0.09	1.10 ± 0.05	99.05 ± 0.26	2.16 ± 0.21	2.88 ± 0.27
	0.05	90.24 ± 0.76	0.44 ± 0.06	0.59 ± 0.05	95.31 ± 0.70	0.80 ± 0.10	1.07 ± 0.07	99.03 ± 0.26	2.21 ± 0.30	2.98 ± 0.32
	0.1	90.11 ± 0.65	0.47 ± 0.07	0.63 ± 0.07	95.27 ± 0.36	0.81 ± 0.12	1.09 ± 0.09	99.05 ± 0.21	2.20 ± 0.28	2.95 ± 0.28
	0.5	89.65 ± 0.71	0.43 ± 0.06	0.60 ± 0.04	94.83 ± 0.80	0.76 ± 0.10	1.05 ± 0.06	99.09 ± 0.18	2.05 ± 0.24	2.84 ± 0.20
	1	89.75 ± 1.03	0.32 ± 0.03	0.69 ± 0.04	94.91 ± 0.63	0.56 ± 0.07	1.20 ± 0.10	99.03 ± 0.29	1.42 ± 0.20	3.08 ± 0.36
	5	90.37 ± 0.43	**0.09 ± 0.03**	**0.33 ± 0.03**	95.67 ± 0.37	**0.14 ± 0.04**	0.51 ± 0.06	99.14 ± 0.17	**0.33 ± 0.08**	1.17 ± 0.08
Pol	10	90.15 ± 0.52	0.15 ± 0.04	0.36 ± 0.02	95.35 ± 0.42	0.19 ± 0.05	**0.47 ± 0.05**	99.19 ± 0.35	0.41 ± 0.16	0.99 ± 0.21
	50	**89.92 ± 0.36**	0.54 ± 0.07	0.72 ± 0.03	94.83 ± 0.62	0.66 ± 0.08	0.88 ± 0.02	99.16 ± 0.18	0.80 ± 0.09	**1.06 ± 0.02**
	100	89.91 ± 0.82	0.91 ± 0.09	1.12 ± 0.05	95.13 ± 0.43	1.05 ± 0.09	1.31 ± 0.04	**99.03 ± 0.23**	1.24 ± 0.11	1.54 ± 0.06
	500	90.21 ± 0.87	2.68 ± 0.23	2.56 ± 0.19	95.13 ± 0.60	2.99 ± 0.20	2.86 ± 0.22	99.10 ± 0.21	3.35 ± 0.18	3.20 ± 0.21
	1000	90.23 ± 0.72	3.56 ± 0.33	2.39 ± 0.41	95.27 ± 0.76	3.79 ± 0.29	2.55 ± 0.47	99.19 ± 0.15	4.07 ± 0.28	2.74 ± 0.52
	5000	90.27 ± 0.61	4.97 ± 0.10	1.43 ± 0.19	95.10 ± 0.52	5.09 ± 0.09	1.47 ± 0.19	98.81 ± 0.20	5.25 ± 0.07	1.52 ± 0.18
	10000	90.41 ± 0.60	5.25 ± 0.11	1.47 ± 0.22	**95.05 ± 0.22**	5.37 ± 0.11	1.51 ± 0.22	98.93 ± 0.16	5.57 ± 0.13	1.56 ± 0.23
	0.01	90.38 ± 0.53	1.70 ± 0.06	2.39 ± 0.23	95.16 ± 0.58	2.66 ± 0.13	3.76 ± 0.45	99.10 ± 0.25	6.01 ± 0.53	8.42 ± 0.64
	0.05	90.28 ± 0.56	1.68 ± 0.08	2.38 ± 0.30	95.12 ± 0.65	2.64 ± 0.17	3.75 ± 0.56	99.13 ± 0.24	6.01 ± 0.35	8.49 ± 0.69
	0.1	90.20 ± 0.46	1.68 ± 0.06	2.34 ± 0.30	95.05 ± 0.48	2.61 ± 0.13	3.65 ± 0.55	99.17 ± 0.23	6.05 ± 0.48	8.39 ± 0.78
	0.5	90.31 ± 0.65	1.64 ± 0.10	2.06 ± 0.34	95.04 ± 0.38	2.50 ± 0.18	3.14 ± 0.55	99.06 ± 0.27	5.35 ± 0.55	6.74 ± 1.39
	1	90.17 ± 0.39	1.54 ± 0.07	1.19 ± 0.16	94.80 ± 0.46	2.24 ± 0.12	1.74 ± 0.25	98.98 ± 0.12	4.42 ± 0.34	3.43 ± 0.55
	5	89.79 ± 0.36	**1.29 ± 0.04**	0.47 ± 0.02	**95.02 ± 0.27**	1.69 ± 0.05	0.62 ± 0.03	98.83 ± 0.16	3.03 ± 0.11	1.11 ± 0.06
Superconduct	10	89.79 ± 0.55	1.39 ± 0.05	0.46 ± 0.04	94.96 ± 0.23	**1.65 ± 0.04**	0.54 ± 0.04	98.85 ± 0.20	2.77 ± 0.10	0.91 ± 0.09
	50	89.84 ± 0.31	1.87 ± 0.03	**0.41 ± 0.02**	94.95 ± 0.53	2.04 ± 0.03	**0.45 ± 0.03**	**98.99 ± 0.20**	**2.53 ± 0.10**	**0.56 ± 0.05**
	100	**89.86 ± 0.28**	2.11 ± 0.03	0.45 ± 0.03	94.92 ± 0.52	2.28 ± 0.03	0.48 ± 0.03	98.97 ± 0.24	2.66 ± 0.05	**0.56 ± 0.05**
	500	89.57 ± 0.57	2.43 ± 0.04	0.59 ± 0.09	94.68 ± 0.75	2.63 ± 0.03	0.64 ± 0.10	98.94 ± 0.22	3.04 ± 0.08	0.74 ± 0.12
	1000	89.47 ± 0.18	2.53 ± 0.04	0.64 ± 0.07	94.62 ± 0.29	2.78 ± 0.07	0.70 ± 0.08	99.10 ± 0.17	3.31 ± 0.10	0.83 ± 0.11
	5000	89.50 ± 0.71	2.79 ± 0.23	0.70 ± 0.14	94.80 ± 0.35	3.09 ± 0.18	0.77 ± 0.15	99.11 ± 0.16	3.70 ± 0.19	0.92 ± 0.17
	10000	89.51 ± 0.50	2.81 ± 0.17	0.71 ± 0.11	94.86 ± 0.42	3.15 ± 0.12	0.80 ± 0.13	99.06 ± 0.29	3.88 ± 0.12	0.99 ± 0.17
	0.01	90.26 ± 1.30	0.40 ± 0.02	0.19 ± 0.03	95.38 ± 0.28	0.56 ± 0.03	0.27 ± 0.05	98.95 ± 0.46	1.23 ± 0.19	0.58 ± 0.08
	0.05	**90.18 ± 1.48**	0.40 ± 0.02	0.19 ± 0.03	95.51 ± 0.42	0.56 ± 0.03	0.27 ± 0.05	98.97 ± 0.47	1.24 ± 0.19	0.59 ± 0.09
	0.1	90.40 ± 1.17	0.41 ± 0.02	0.20 ± 0.03	95.45 ± 0.38	0.56 ± 0.03	0.27 ± 0.04	98.95 ± 0.50	1.24 ± 0.18	0.59 ± 0.08
	0.5	90.55 ± 1.00	0.41 ± 0.01	0.20 ± 0.02	**95.14 ± 0.36**	0.55 ± 0.01	0.27 ± 0.03	**98.98 ± 0.31**	1.25 ± 0.04	0.61 ± 0.05
	1	90.32 ± 0.87	0.39 ± 0.01	0.16 ± 0.03	94.83 ± 0.36	0.52 ± 0.02	0.22 ± 0.05	99.05 ± 0.26	1.13 ± 0.10	0.47 ± 0.12
	5	90.48 ± 0.85	**0.35 ± 0.01**	0.04 ± 0.01	95.31 ± 0.80	**0.43 ± 0.01**	0.05 ± 0.01	99.02 ± 0.33	0.65 ± 0.05	0.07 ± 0.01
Wine Quality	10	90.46 ± 0.69	**0.35 ± 0.01**	**0.03 ± 0.00**	95.31 ± 0.87	0.43 ± 0.02	**0.04 ± 0.01**	99.03 ± 0.36	**0.63 ± 0.04**	**0.06 ± 0.01**
	50	90.92 ± 0.59	0.46 ± 0.03	0.05 ± 0.01	95.52 ± 0.66	0.54 ± 0.04	0.06 ± 0.01	99.08 ± 0.19	0.73 ± 0.05	0.09 ± 0.01
	100	90.57 ± 0.62	0.56 ± 0.06	0.07 ± 0.01	95.38 ± 0.62	0.65 ± 0.05	0.09 ± 0.01	99.23 ± 0.31	0.83 ± 0.05	0.11 ± 0.01
	500	90.45 ± 0.81	1.04 ± 0.16	0.16 ± 0.02	95.31 ± 0.88	1.18 ± 0.19	0.18 ± 0.03	99.12 ± 0.57	1.45 ± 0.23	0.22 ± 0.03
	1000	90.69 ± 1.27	1.27 ± 0.18	0.20 ± 0.04	95.77 ± 1.02	1.45 ± 0.22	0.23 ± 0.04	99.11 ± 0.57	1.76 ± 0.25	0.27 ± 0.05
	5000	90.26 ± 1.64	1.72 ± 0.35	0.33 ± 0.10	95.60 ± 1.22	1.97 ± 0.35	0.38 ± 0.12	99.25 ± 0.30	2.34 ± 0.36	0.45 ± 0.14
	10000	90.57 ± 1.60	2.19 ± 0.72	0.52 ± 0.21	95.54 ± 1.22	2.44 ± 0.69	0.58 ± 0.22	99.28 ± 0.30	2.85 ± 0.72	0.68 ± 0.24
	0.01	89.16 ± 0.64	85.77 ± 13.72	340.83 ± 62.37	94.87 ± 0.29	129.24 ± 27.41	511.50 ± 109.12	98.98 ± 0.13	231.64 ± 66.53	915.42 ± 268.53
	0.05	89.86 ± 0.80	128.65 ± 55.28	452.92 ± 139.28	95.09 ± 0.54	195.72 ± 88.58	683.63 ± 226.99	**99.00 ± 0.20**	353.86 ± 174.01	1225.39 ± 456.33
	0.1	89.97 ± 0.69	66.60 ± 27.49	270.27 ± 102.54	95.07 ± 0.38	101.60 ± 44.58	412.20 ± 166.88	99.18 ± 0.22	201.74 ± 91.76	819.73 ± 346.97
	0.5	90.13 ± 0.35	64.62 ± 21.31	204.61 ± 59.75	94.94 ± 0.44	95.03 ± 33.34	300.99 ± 94.93	99.12 ± 0.21	179.22 ± 59.02	570.41 ± 173.99
	1	**90.00 ± 0.70**	79.64 ± 10.16	141.85 ± 19.99	95.04 ± 0.42	148.52 ± 17.69	263.94 ± 30.69	98.92 ± 0.23	377.03 ± 27.50	670.45 ± 57.30
	5	90.03 ± 1.29	**22.51 ± 2.28**	**13.04 ± 3.21**	95.17 ± 0.71	27.59 ± 2.98	**16.02 ± 4.19**	99.27 ± 0.28	41.19 ± 5.90	24.02 ± 7.20
Drift	10	89.89 ± 0.67	22.08 ± 3.21	12.95 ± 4.35	95.27 ± 0.52	26.96 ± 3.11	15.79 ± 4.99	98.92 ± 0.35	38.15 ± 3.36	**22.30 ± 6.78**
	50	89.62 ± 0.59	25.20 ± 12.18	13.69 ± 8.50	94.87 ± 0.59	31.00 ± 15.03	16.85 ± 10.49	99.02 ± 0.40	44.21 ± 20.26	23.92 ± 14.24
	100	89.99 ± 1.25	22.86 ± 1.62	14.40 ± 3.94	**94.98 ± 0.85**	**27.40 ± 1.81**	17.28 ± 4.80	98.84 ± 0.22	**37.37 ± 3.88**	23.71 ± 7.60
	500	89.98 ± 0.70	31.85 ± 4.66	19.99 ± 5.62	95.11 ± 1.16	38.08 ± 5.74	23.90 ± 6.74	99.05 ± 0.45	51.99 ± 8.98	32.47 ± 9.00
	1000	90.10 ± 0.96	37.36 ± 2.79	24.70 ± 5.41	95.27 ± 0.51	43.06 ± 2.86	28.43 ± 5.91	99.14 ± 0.24	54.99 ± 1.66	36.29 ± 7.02
	5000	90.24 ± 1.13	58.13 ± 13.22	23.56 ± 8.94	94.87 ± 0.55	67.31 ± 17.24	27.31 ± 11.00	99.16 ± 0.31	81.31 ± 18.65	33.07 ± 12.80
	10000	90.19 ± 0.82	82.49 ± 12.06	26.64 ± 15.71	95.29 ± 0.57	99.55 ± 14.85	31.97 ± 18.65	99.20 ± 0.05	139.84 ± 34.11	43.91 ± 26.11



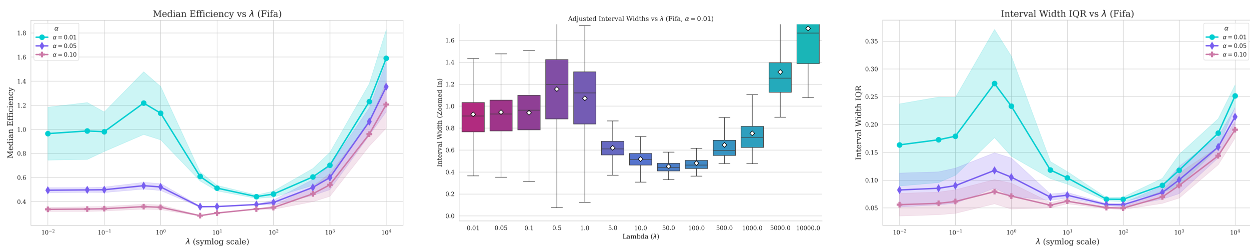
(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 21: Results analysis for varying λ values for the California dataset.



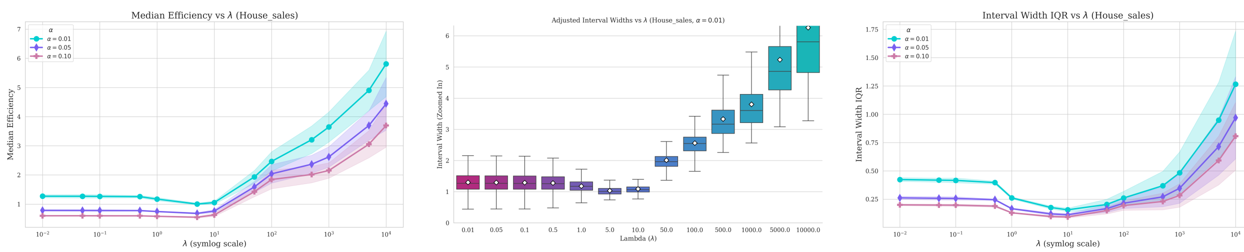
(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 22: Results analysis for varying λ values for the Diamonds dataset.



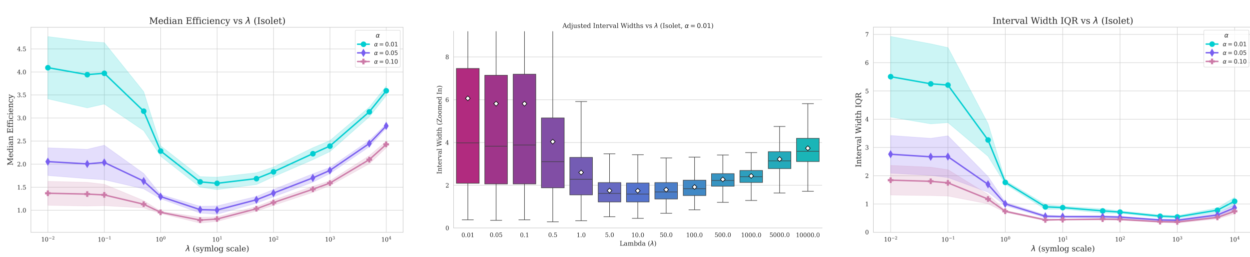
(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 23: Results analysis for varying λ values for the Fifa dataset.



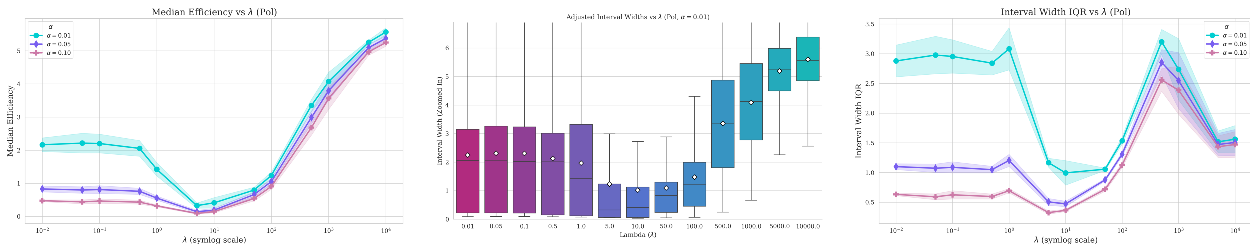
(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 24: Results analysis for varying λ values for the House Sales dataset.



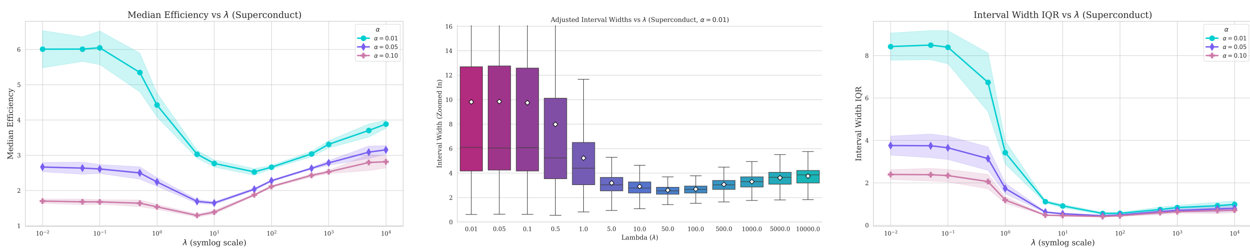
(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 25: Results analysis for varying λ values for the Isolet dataset.



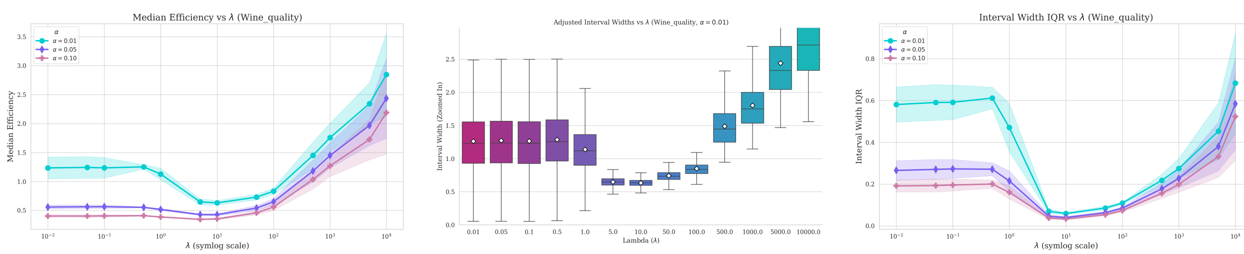
(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 26: Results analysis for varying λ values for the Pol dataset.



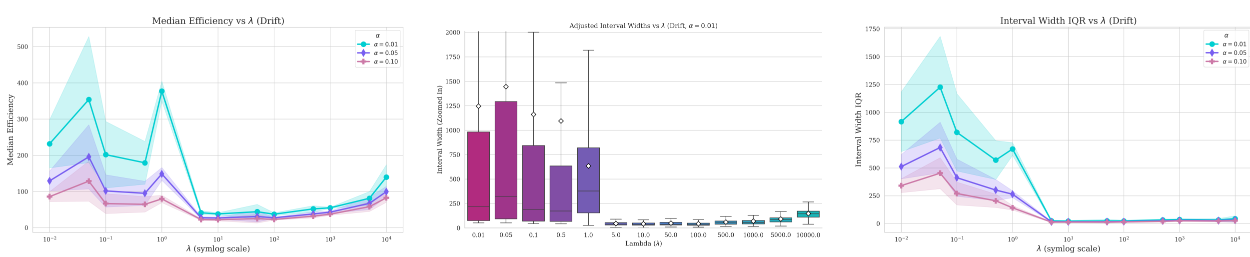
(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 27: Results analysis for varying λ values for the Superconduct dataset.



(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 28: Results analysis for varying λ values for the Wine Quality dataset.



(a) Median Efficiency. (b) Width Boxplots. (c) Interquartile Range (IQR).

Figure 29: Results analysis for varying λ values for the Drift dataset.