Statistica Sinica Preprint No: SS-2025-0383	
Title	Efficient Prediction Intervals Via Debiased Conformal Threshold Ridge Regression
Manuscript ID	SS-2025-0383
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0383
Complete List of Authors	Jiamei Wu, Pan Shang, Yanlin Tang, Linglong Kong, Bei Jiang and Lingchen Kong
Corresponding Authors	Linglong Kong
E-mails	lkong@ualberta.ca
Notice: Accepted author version.	

Statistica Sinica

Efficient Prediction Intervals via Debiased Conformal Threshold Ridge Regression

Jiamei Wu[1], Pan Shang[1], Yanlin Tang[2], Linglong Kong[3], Bei Jiang[3] and Lingchen Kong[1]

[1]*School of Mathematical and Statistics, Beijing Jiaotong University*

[2] *KLATASDS-MOE, School of Statistics, East China Normal University*

[3] *Department of Mathematical and Statistical Sciences, University of Alberta*

Abstract: Uncertainty quantification is a critical aspect of modern statistical modeling and machine learning. Among many methods for uncertainty quantification, conformal prediction is a powerful one, which offers finite-sample coverage guarantees under the weak assumption of exchangeability. However, the efficiency of conformal prediction in high-dimensional settings is often compromised by the overfitting of complex models or the inherent bias. To address this, we propose the debiased conformal threshold ridge regression (DeCThRR), a computationally efficient framework that integrates a stable thresholded ridge regression estimator with a targeted procedure to correct for regularization-induced bias, before computing nonconformity scores. We prove that our method preserves finite-sample marginal coverage while achieving near-optimal efficiency and asymptotic conditional coverage under mild assumptions. Experiments confirm that our method produces narrower, more reliable prediction intervals than standard conformal approaches and some advanced inference methods, demonstrating

remarkable robustness even under model misspecification.

Key words and phrases: Conformal prediction, High-dimensional regression, Bias correction, Threshold ridge regression, Finite-sample coverage, Asymptotic efficiency, Conditional coverage

1. Introduction

Uncertainty quantification (UQ) is crucial in statistical modeling and machine learning, particularly in contexts involving high-stakes decision-making in fields such as medicine (Bashari et al., 2023; Lu et al., 2022; Vazquez and Facelli, 2022) and finance (Chernozhukov et al., 2021). In regression analysis, constructing reliable prediction intervals is paramount for quantifying predictive risk and supporting downstream decisions.

Existing approaches to uncertainty quantification can be broadly divided into parameter-level methods and prediction-level methods, along with other related techniques. Parameter-level methods include post-selection inference (Liu and Yu, 2013; Berk et al., 2013; Lee et al., 2016; Tibshirani et al., 2018; Zrnic and Jordan, 2023) and debiasing approaches (Javanmard and Montanari, 2014; Van de Geer et al., 2014; Zhang and Politis, 2022), while prediction-level methods encompass bootstrap techniques (Efron and Tibshirani, 1994; Mammen, 1993; Chetverikov and Kato, 2013) and confor-

mal prediction (CP) (Vovk et al., 2005; Lei et al., 2018; Fontana et al., 2023). While advanced methods such as post-selection inference and debiasing are powerful, they typically rely on assumptions of Gaussianity or specific sparsity structures that may not hold in practice. Moreover, these approaches primarily quantify uncertainty for model coefficients rather than for predictions. Bootstrap methods provide a general framework for predictive UQ but offer only asymptotic validity. In contrast, CP has emerged as a powerful framework for predictive uncertainty quantification. It requires only exchangeability—a condition weaker than the i.i.d. assumption—and makes no Gaussianity or sparsity assumptions. Most importantly, CP yields prediction intervals with finite-sample coverage guarantees, a distinct advantage over methods with merely asymptotic guarantees. Its model-agnostic nature further enables integration with virtually any predictive algorithm, making it a highly flexible tool for UQ.

In high-dimensional settings, CP is typically implemented by wrapping a base predictor with split conformal calibration, which preserves finite-sample marginal coverage but makes interval informativeness depend on the predictor and the modeling pipeline. Representative high-dimensional conformal constructions build on regularized linear estimators and efficient implementations (Hebiri, 2010; Burnaev and Vovk, 2014; Lei, 2019), with

recent advances on full conformal theory in proportional regimes (Gibbs and Candès, 2025) and scalable AMP-based computation for high-dimensional GLMs (Clarté and Zdeborová, 2025), as well as calibration after efficiency-oriented model selection (Liang et al., 2024). While exchangeability underpins the finite-sample validity of classical conformal prediction, it is often violated in practice. Motivated by this, weighted conformal methods address covariate shift (Tibshirani et al., 2019) and broader procedures relax exchangeability under drift or online shifts (Barber et al., 2023). In high-dimensional covariate shift where likelihood-ratio estimation is difficult, Joshi et al. (2025) propose likelihood-ratio regularization. Practically, flexible learners such as random forests or neural networks may overfit and inflate nonconformity scores, yielding wide intervals (Angelopoulos and Bates, 2023), whereas ridge-type stabilization can shorten intervals but may introduce shrinkage bias, see Cases 1–4. This stability–bias trade-off motivates our approach: we correct regularization-induced bias before computing nonconformity scores, enabling threshold ridge regression to retain computational and stability advantages while producing more informative conformal intervals.

Based on the above considerations, we propose debiased conformal threshold ridge regression (DeCThRR), a computationally efficient frame-

work that integrates a stable ThRR estimator with a targeted procedure to correct for regularization-induced bias when computing nonconformity scores. Our approach explicitly incorporates debiasing corrections into the nonconformity scores, thereby enhancing predictive accuracy while preserving the distribution-free validity inherent in CP. We theoretically demonstrate that the prediction intervals of DeCThRR satisfy finite-sample marginal coverage. Furthermore, under specified regularity conditions, these intervals possess near-optimal efficiency and achieve asymptotic conditional coverage. Numerical experiments reveal that this framework is remarkably robust. Even when the underlying linear model is misspecified, it often produces more efficient and reliable intervals than complex, non-linear models that are prone to overfitting in high-dimensional settings.

The paper is organized as follows. Section 2 gives a brief introduction of CP. In Section 3, we propose the debiased conformal prediction, apply it to the threshold ridge regression framework, and present the corresponding theoretical results. Extensive numerical experiments demonstrate that our method systematically outperforms its non-debiased counterpart, standard conformal methods based on the Lasso, random forests, and neural networks, as well as post-selection strategies and bootstrap in Section 4. Section 5 contains further remarks and future directions.

2. Preliminaries

This section lays the theoretical groundwork for our proposed methodology. We begin by reviewing the basic idea of conformal prediction (CP), which provides the distribution-free guarantees for uncertainty quantification. Subsequently, we introduce threshold ridge regression, the computationally efficient high-dimensional model that serves as the base estimator in our approach. Finally, we discuss the impact of estimation bias in regularized estimators on the length of prediction interval.

2.1 Conformal Prediction

Let $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}, i = 1, ..., n$ denote training samples. Given a desired miscoverage rate $\alpha \in (0, 1)$, CP constructs a prediction band $\widehat{C} : \mathbb{R}^p \to \mathcal{Y} \subseteq \mathbb{R}$ for Y_{n+1} at a test point X_{n+1} satsifying $\mathbb{P}(Y_{n+1} \in \widehat{C}(X_{n+1})) \geq 1-\alpha$, under the assumption that all pairs $(X_i, Y_i)_{i=1}^{n+1}$ are exchangeable. A sequence of random variables is defined as exchangeable if its joint probability distribution is invariant under any permutation of its indices. Since CP only relies on the assumption of exchangeability, it is a flexible approach that can be applied using various algorithms, including supervised settings such as regression and classification, and unsupervised settings such as clustering and principal components analysis.

In practice, full CP is computationally intensive because it requires model refitting for each candidate $y \in \mathbb{R}$, which is particularly prohibitive in regression. A widely used and computationally feasible alternative is split conformal prediction (SCP) (Lei et al., 2018), which partitions the data into a training set D_{train} and a calibration set D_{cal} of sizes n_{train} and n_{cal}, respectively. This avoids repeated model retraining while preserving distribution-free, finite-sample coverage guarantees. The coverage property is stated in Proposition 1, and the procedure is summarized in Algorithm 1.

Proposition 1 (Vovk et al., 2005; Lei et al., 2018)**.** *Assume that the data pairs* $(X_1, Y_1), \ldots, (X_n, Y_n)$, (X_{new}, Y_{new}) *are exchangeable. Let* $\widehat{C}(X_{new})$ *be the prediction interval constructed by Algorithm 1. Then for any* $\alpha \in (0, 1)$*,*

$$\mathbb{P}(Y_{new} \in \widehat{C}(X_{new})) \geq 1 - \alpha.$$

Furthermore, if the distribution of the nonconformity score $R_i = |Y_i - \hat{\mu}(X_i)|$ *is continuous, then*

$$\mathbb{P}(Y_{new} \in \widehat{C}(X_{new})) \leq 1 - \alpha + \frac{1}{n_{cal} + 1}.$$

2.2 Threshold Ridge Regression

Our proposed method builds upon threshold ridge regression (Shao and Deng, 2012). We select this model as our base estimator for its robustness to

Algorithm 1 Split Conformal Prediction

Input:

Data $(X_i, Y_i), i = 1, ..., n$, prescribed error level α, underlying model μ, points $X_{\text{new}} = \{X_{n+1}, X_{n+2}, ...\}$ which are to construct prediction bands.

Output:

Prediction bands at each point in $\mathcal{X}_{\text{new}}$.

1: **for** $X \in \mathcal{X}_{\text{new}}$ **do**

2: Randomly split the dataset into two disjoint subsets D_{train} and D_{cal};

3: Fit model $\widehat{\mu}$ on the training set D_{train};

4: Calculate nonconformity scores on calibration set

$$R_i = |Y_i - \widehat{\mu}(X_i)|, \quad (X_i, Y_i) \in D_{\text{cal}};$$

5: Sort $R_1, ..., R_{n_{\text{cal}}}$ in the ascending order obtaining $R_{(1)}, ..., R_{(n_{\text{cal}})}$;

6: Construct prediction interval

$$\widehat{C}(X) = [\widehat{\mu}(X) - R_{(\lceil (1-\alpha)(n_{\text{cal}}+1) \rceil)}, \widehat{\mu}(X) + R_{(\lceil (1-\alpha)(n_{\text{cal}}+1) \rceil)}];$$

7: **end for**

Return: $\widehat{C}(X)$ for each point in $\mathcal{X}_{\text{new}}$.

multicollinearity—a common challenge in high-dimensional settings where methods like Lasso can be unstable—and its computational simplicity stemming from a closed-form solution.

First, consider the linear model

$$Y = \mathbf{X}\beta + \epsilon, \tag{2.1}$$

where $Y \in \mathbb{R}^n$ is the response vector, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the design matrix, $\beta \in \mathbb{R}^p$ is the vector of unknown regression coefficients, and $\epsilon \in \mathbb{R}^n$ is the random error vector. Perform a thin singular value decomposition (SVD), which refers to the reduced SVD keeping only the r nonzero singular values and their corresponding singular vectors; see more details in Theorem 7.3.2 in (Horn and Johnson, 2012). Specifically, $\mathbf{X} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}^\top$, where $\mathbf{P}$ and $\mathbf{Q}$ are $n\times r$ and $p\times r$ orthonormal matrices, $\mathbf{\Lambda}$ is an $r\times r$ diagonal matrix, r denotes the given rank. Denote $\mathbf{Q}_\perp$ as the $p \times (p-r)$ orthonormal complement of $\mathbf{Q}$, which satisfies the following properties

$$\mathbf{Q}_\perp^\top \mathbf{Q}_\perp = \mathbf{I}_{p-r}, \quad \mathbf{Q}^\top \mathbf{Q}_\perp = 0 \quad \text{and} \quad \mathbf{Q}\mathbf{Q}^\top + \mathbf{Q}_\perp \mathbf{Q}_\perp^\top = \mathbf{I}_p.$$

Define $\theta = \mathbf{Q}\mathbf{Q}^\top\beta$ and $\theta_\perp = \mathbf{Q}_\perp\mathbf{Q}_\perp^\top\beta$, so that $\beta = \theta + \theta_\perp$. According to Shao and Deng (2012), the ridge regression is used to estimate θ rather than β directly, where the estimator in ridge regression is

$$\widehat{\theta} = \left(\mathbf{X}^\top\mathbf{X} + h_n\mathbf{I}_p\right)^{-1}\mathbf{X}^\top y, \tag{2.2}$$

where $h_n > 0$ is a tuning parameter and $\mathbf{I}_p$ denotes the p-dimensional identity matrix. While this distinction matters for coefficient inference, estimating θ is enough for inference about parameters $\boldsymbol{X}\beta = \boldsymbol{X}\theta$ and prediction. ThRR introduces a hard-thresholding step to the ridge estimates. The ThRR estimator $\widetilde{\theta}$, is defined as

$$\widetilde{\theta}_j = \widehat{\theta}_j \cdot \mathbf{1}(|\widehat{\theta}_j| > a_n), \quad \text{for } j = 1, \ldots, p, \tag{2.3}$$

where $\widehat{\theta}_j$ is the j-th component of the ridge regression estimator, $a_n > 0$ is the thresholding parameter, and $\mathbf{1}(\cdot)$ is the indicator function. This post-processing step effectively sets small ridge coefficients to zero, thereby it induces sparsity. It offers a computationally simple approach with a closed-form solution, making it attractive for high-dimensional settings. Its ability to handle multicollinearity and provide a stable solution, combined with its sparsity-inducing mechanism, makes it a suitable candidate as a base estimator in high-dimensional contexts.

2.3 Bias in High-Dimensional Settings

First, we conduct a small simulation study in the one-dimensional setting (Figure 1) to illustrate how correcting for estimation bias improves the performance of standard conformal prediction. After bias correction, the resulting prediction interval length is much closer to that of the oracle

interval. However, most traditional estimators are inevitably biased, particularly in high-dimensional settings, since a point estimate $\widehat{\theta} \in \mathbb{R}^p$ must be produced from data in lower dimension. For a more detailed discussion, see Zhang and Zhang (2014) and Javanmard and Montanari (2014). Taking ridge regression as an example, suppose the parameter of interest is $a^\top \beta$ in the linear model $y = \mathbf{X}\beta + \epsilon$, where $\mathbb{E}(\epsilon) = \mathbf{0}$, $\mathrm{Var}(\epsilon) = \sigma^2 \mathbf{I}_n$, and $\mathbf{a}$ is a given vector. This bounded-variance assumption is made solely to obtain explicit bias–variance formulas (2.4) and (2.5). This assumption is local to this subsection and is not used in any coverage or efficiency theorem. The ridge estimator is $a^\top \widehat{\theta}$ with $\widehat{\theta} = \left(\mathbf{X}^\top \mathbf{X} + h_n \mathbf{I}_p\right)^{-1} \mathbf{X}^\top y$, for some $h_n > 0$. Assume that the error vector $\boldsymbol{\epsilon}$ consists of i.i.d. components. Then the bias and the standard deviation can be calculated as follows,

$$\mathbb{E} a^\top \widehat{\theta} - a^\top \beta = -h_n a^\top (\mathbf{X}^\top \mathbf{X} + h_n \mathbf{I}_p)^{-1} \beta,$$

which implies

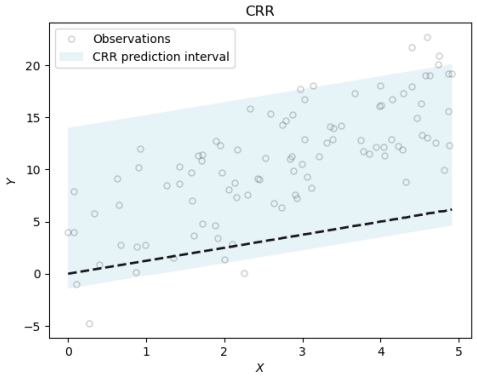
$$\left| \mathbb{E} a^\top \widehat{\theta} - a^\top \beta \right| \leq \frac{h_n \|a\|_2 \times \|\beta\|_2}{\lambda_{\min} + h_n}, \tag{2.4}$$

$$\begin{aligned} \sqrt{\mathrm{Var}\left(a^\top \widehat{\theta}\right)} &= \sqrt{\mathrm{Var}\left(\epsilon_1\right) \times a^\top (\mathbf{X}^\top \mathbf{X} + h_n \mathbf{I}_p)^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X} + h_n \mathbf{I}_p)^{-1} a} \\ &\leq \frac{\sqrt{\mathrm{Var}\left(\epsilon_1\right)} \times \|a\|_2}{\sqrt{\lambda_{\min}}}, \end{aligned} \tag{2.5}$$

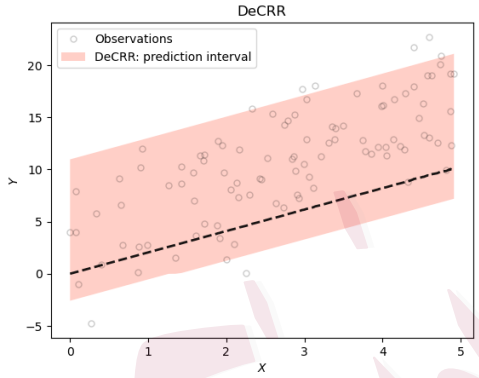
where $\lambda_{\min}$ is the smallest singular value of $\mathbf{X}^\top \mathbf{X}$. If $\|\beta\|_2$ does not have a bounded order, the bias is significantly larger than the standard devia-

tion, which complicates the construction of prediction intervals (Zhang and Politis, 2022). Such a large bias systematically displaces the center of the prediction interval from the true value. Consequently, an interval will be unreliable for inference as its true coverage probability is substantially lower than its nominal level. Although CP guarantees the nominal coverage, the resulting intervals are so wide that decreases practical informational value.

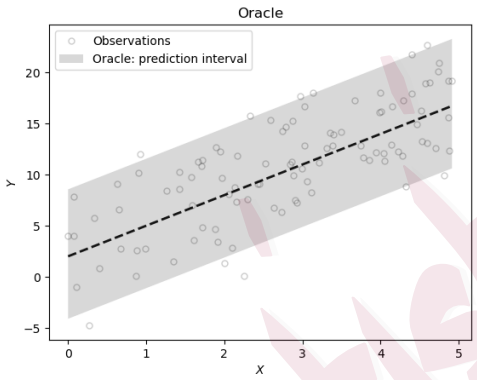
Recall the split conformal prediction band constructed in Algorithm 1. The width of band is $2R_{(\lceil(1-\alpha/2)(n_{\text{cal}}+1)\rceil)}$. A natural intuition is that we can reduce bias when constructing the nonconformity score function and then the resulting prediction interval may be more efficient.



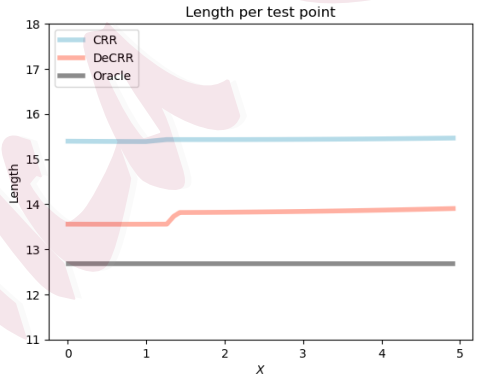
(a) Conformalized ridge regression (CRR): Avg. coverage 90%; Avg. length 14.79.



(b) Debiased conformalized ridge regression(DeCRR): Avg. coverage 91%; Avg. length 13.40.



(c) Oracle prediction interval: Avg. coverage 90%; Avg. length 12.90.



(d) Length of predictions intervals.

Figure 1: Prediction intervals on simulated data with outliers: (a) the standard conformalized ridge regression, (b) conformalized ridge regression after debiasing, and (c) oracle prediction interval. The length of the interval as a function of X is shown in (d). The target coverage rate is 90%. The broken black curve in (a), (b) and (c) is the pointwise prediction from the ridge regression.

3. Debiased Conformal Threshold Ridge Regression

To resolve the dilemma between the overfitting of complex models and the estimation bias of traditional regularized methods, this section introduces our proposed solution: debiased conformal thresholded ridge regression (DeCThRR). It produces prediction intervals that are not only distribution-free and valid in finite samples, but are also substantially narrower theoretically.

3.1 DeCThRR

CP has emerged as a uniquely powerful framework for uncertainty quantification, prized for its ability to provide distribution-free prediction intervals with guaranteed finite-sample coverage. However, its practical utility in high-dimensional settings is hampered by a fundamental dilemma in the choice of the underlying algorithm. Complex models like neural networks or random forests are prone to overfitting, which inflates nonconformity scores and yields intervals that are valid but too wide to be informative. Therefore, we consider threshold ridge regression as underlying method, which is stable and computationally efficient. But it introduces systematic regularization bias that similarly contaminates the scores and degrades interval efficiency. To handle this, we account for the bias when constructing the

nonconformity score function, i.e.,

$$\widetilde{R}_i = |Y_i - \widehat{\mu}(X_i) + \text{bias}(\widehat{\mu}(X_i))|, \quad i = 1, ..., n_{\text{cal}}, \tag{3.6}$$

where the bias is estimated by training data. Then the prediction band is

$$\widehat{C}_n^{\text{Debias}}(X_{n+1}) = \left\{ y \in \mathbb{R} : \widetilde{R}_{n+1} \leq \text{Q}_{1-\alpha}\left(\sum_{i=1}^{n_{\text{cal}}} \frac{1}{n_{\text{cal}}+1} \cdot \delta_{\widetilde{R}_i} \right) \right\},$$

where $\text{Q}_{1-\alpha}(\cdot)$ is empirical $(1-\alpha)$-quantile function, and $\delta_{\widetilde{R}_i}$ represents a point mass at the location of the i-th nonconformity score $\widetilde{R}_i$.

Consider the regression model in (2.1). For a threshold a_n, we define the set as $\mathcal{M}_{a_n} = \left\{ j \mid |\widetilde{\theta}_j| > a_n \right\}$, $\widetilde{\theta}_j = \widehat{\theta}_j \times \mathbf{1}_{j \in \mathcal{M}_{a_n}}$. Let g_n denote the number of elements in the set $\mathcal{M}_{a_n}$. We define $c_{ik} = \sum_{j \in \mathcal{M}_{a_n}} x_{ij} q_{jk}, \forall i = 1, ..., n, k = 1, ..., r$, where q_{jk} denotes the (j,k)-th entry of $\mathbf{Q}$. To mitigate the bias of the estimator, we define the debiased nonconformity score function as

$$\widetilde{R}_i = |Y_i - X_i\widetilde{\theta} - h_n X_i \boldsymbol{Q}(\boldsymbol{\Lambda}^2 + h_n \mathbf{I}_r)^{-1}\mathbf{Q}^\top \widetilde{\theta}|, \quad i \in D_{\text{cal}}, \tag{3.7}$$

where $-h_n X_i \boldsymbol{Q}(\boldsymbol{\Lambda}^2 + h_n \mathbf{I}_r)^{-1}\mathbf{Q}^\top \widetilde{\theta}$ is estimation of the bias. The prediction band $\widehat{C}_{\text{DeCThRR}}(X_{n+1})$ is

$$\widehat{C}_{\text{DeCThRR}}(X_{n+1}) = [X_{n+1}\widetilde{\theta} + \widehat{b}(X_{n+1}) - \widetilde{R}_k, X_{n+1}\widetilde{\theta} + \widehat{b}(X_{n+1}) + \widetilde{R}_k],$$

where $\widehat{b}(X) = h_n X \mathbf{Q}(\boldsymbol{\Lambda}^2 + h_n \mathbf{I}_r)^{-1}\mathbf{Q}^\top \widetilde{\theta}$ and $\widetilde{R}_k$ is the $\lceil (1-\alpha)(n_{\text{cal}}+1) \rceil$-th smallest one among $\{\widetilde{R}_1, ..., \widetilde{R}_{n_{\text{cal}}}\}$. The procedure is summarized in Algorithm 2.

Algorithm 2 DeCThRR

Input:

Data $(X_i, Y_i), i = 1, ..., n$, prescribed error level α, threshold parameter a_n and ridge parameter h_n, points $\mathcal{X}_{\text{new}} = \{X_{n+1}, X_{n+2}, ...\}$ which are to construct prediction bands

Output:

Prediction bands at each point in $\mathcal{X}_{\text{new}}$

1: **for** $X \in \mathcal{X}_{\text{new}}$ **do**

2: Randomly split the dataset into two disjoint subsets D_{train} and D_{cal};

3: Estimate $\widetilde{\theta}$ and the bias on the training set D_{train};

4: Calculate debiased nonconformity score on calibration set as

$$\widetilde{R}_i = |Y_i - X_i\widetilde{\theta} - h_n X_i \boldsymbol{Q}(\boldsymbol{\Lambda}^2 + h_n \mathbf{I}_r)^{-1}\mathbf{Q}^\top\widetilde{\theta}|, \quad (X_i, Y_i) \in D_{\text{cal}};$$

5: Sort $\widetilde{R}_1, ..., \widetilde{R}_{n_{\text{cal}}}$ in the ascending order obtaining $\widetilde{R}_{(1)}, ..., \widetilde{R}_{(n_{\text{cal}})}$;

6: Calculate $\widehat{b}(X)$ and construct the prediction interval

7: **end for** $\widehat{C}_{\text{DeCThRR}}(X) = [X\widetilde{\theta} + \widehat{b}(X) - \widetilde{R}_k, X\widetilde{\theta} + \widehat{b}(X) + \widetilde{R}_k]$;

Return: $\widehat{C}_{\text{DeCThRR}}(X)$ for each point in $\mathcal{X}_{\text{new}}$.

3.2 Theoretical Properties

The following result shows that the debiased conformal prediction bands retain finite sample validity.

Theorem 1. *Suppose* $(X_i, Y_i), i = 1, \ldots, n$ *are exchangeable. Then for a new exchangeable pair* (X_{n+1}, Y_{n+1})*, we have*

$$\mathbb{P}\left(Y_{n+1} \in \widehat{C}_{DeCThRR}\left(X_{n+1}\right)\right) \geq 1 - \alpha.$$

If the fitted absolute residuals $\{\widetilde{R}_i\}_{i=1}^{n+1}$ *are assumed addtionally having continuous joint distribution for all* $y \in \mathbb{R}$*, then we have*

$$\mathbb{P}\left(Y_{n+1} \in \widehat{C}_{DeCThRR}(X_{n+1})\right) \leq 1 - \alpha + \frac{1}{n_{cal} + 1}.$$

The bias-correction step of the nonconformity score function remains the exchangeability. Therefore, the proof of this theorem is similar to the classical conformal prediction.

To analyze the efficiency of the prediction band, we first collect some common assumptions that will be used throughout this paper. Further assumptions will be stated when they are needed.

A1 We observe i.i.d data $(X_i, Y_i), i = 1, \ldots, n+1$ from a common distribution P on $\mathbb{R}^p \times \mathbb{R}$ with mean function $\mu(x) = \mathbb{E}(Y|X = x)$.

A2 For (X_i, Y_i), the noise variable $\epsilon_i = Y_i - \mu(X_i)$ is independent of X_i, and the density function of ϵ_i is symmetric about 0 and nonincreasing on $\mathbb{R}_+$.

A3 The density function of $|\epsilon_i|$ is bounded away from zero by $\underline{f} > 0$ in

a neighborhood of its α upper quantile, and is bounded above across its entire support by a constant $\overline{f} < \infty$.

Assumption A1 is a common assumption in regression literature (Zhang and Zhang, 2014; Lei and Wasserman, 2014; Zhang and Politis, 2022). Assumption A2 is relatively mild compared to conditions used in prior work (Zhang and Politis, 2022). It imposes symmetry and monotonicity of the noise density, without requiring any moment assumptions. This assumption is invoked solely to establish the optimality of the oracle prediction band (3.8). Assumption A3 is crucial for ensuring that the estimator of the α upper quantile is close to its true value, which is essential for the proof. Specifically, the quantile function of ϵ_i satisfies γ-Hölder continuity at its α upper quantile with $\gamma = 1$; see Lemma 1 in the supplementary materials.

Inspired by Lei et al. (2018), to quantify the efficiency of the prediction bands, we compare its length to the idealized prediction band. Our theoretical work focuses on the linear regression model, where we denote $\mu(x) = x^\top \beta$ with the parameter vector $\beta \in \mathbb{R}^p$. The oracle prediction band is defined as

$$C_o^*(x) = [\mu(x) - q_{1-\alpha}, \mu(x) + q_{1-\alpha}], \tag{3.8}$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of the distribution of the absolute error $|\epsilon|$. This band assumes complete knowledge of the regression function $\mu(x)$

and the error distribution. Under Assumptions A1 and A2, the band is optimal in the sense outlined in Lei et al. (2018) as follows.

- It has the valid conditional coverage: $\mathbb{P}(Y \in C_o^*(x) \mid X = x) \geq 1 - \alpha$.
- It has the shortest length among all bands with conditional coverage.
- It has the shortest average length among all bands with marginal coverage.

Besides the assumptions above, we require some additional high-dimensional regularity assumptions, which are collected in Section S1 of the supplementary materials. To keep the main text concise, we only summarize their role here. In particular, Theorems 2–4 are established conditional on the realized training design matrix $\mathbf{X}_{\text{train}}$. In this conditional formulation, the ambient dimension p is allowed to diverge with n under the polynomial growth condition $p = O(n^{\alpha_p})$ in Assumption B2, and its effect enters the theory implicitly through the regularity conditions in Section S1 rather than appearing explicitly in the final rate expressions. Taken together, even if some assumptions are violated, our prediction intervals still maintain marginal validity, and the numerical results suggest that they may remain efficient even under model misspecification.

To motivate the need for debiasing, we begin by examining the effi-

ciency of conformal prediction when threshold ridge regression serves as the underlying estimator. The next result formalizes the efficiency of the prediction intervals generated by the standard conformalized threshold ridge regression.

Theorem 2. *Fix $\alpha \in (0,1)$. Assume the linear model is correctly specified. Let $\widehat{C}_{CThRR}(\cdot)$ denote the conformal prediction interval of the threshold ridge regression. Under Assumptions A1, A3 and B1-B6, we have*

$$\text{Width}\left(\widehat{C}_{CThRR}(X_{n+1})\right) - 2q_{1-\alpha} = O_p(n^{\alpha_\theta - 2\eta + \delta}). \tag{3.9}$$

Here, η denotes the growth exponent of the smallest positive singular value of $\mathbf{X}_{\text{train}}$, α_θ characterizes the growth of the signal norm $\|\theta\|$, and δ is the rate exponent of the ridge parameter h_n. The restriction on δ ensures that the regularization bias remains asymptotically controlled. Further details are given in Section S1 of the Supplementary Material. Theorem 2 demonstrates that the conformal interval converges to the oracle prediction band if $\alpha_\theta < 2\eta - \delta$ as $n \to \infty$. Unfortunately, α_θ is typically greater than $2\eta - \delta$, and thus the efficiency of the conformal interval generally lacks theoretical guarantees. The efficiency of the intervals generated by debiased conformal prediction on the threshold ridge regression is outlined as follows.

Theorem 3. *Fix $\alpha \in (0,1)$. Assume the linear model is correctly specified.*

Let $\widehat{C}_{DeCThRR}(\cdot)$ denote the conformal prediction interval through debiasing in the threshold ridge regression. Under Assumptions A1, A3 and B1-B6 as in Theorem 2, we have

$$\text{Width}\left(\widehat{C}_{DeCThRR}(X_{n+1})\right) - 2q_{1-\alpha} = O_p(n^{-\eta}). \tag{3.10}$$

Since η is usually positive, the interval produced by our method converges to the oracle prediction interval in Lei et al. (2018) at a certain rate, whereas the classical conformal prediction interval may not. The proof of Theorem 3 is presented in the supplementary materials.

Assumption A2 is used only to motivate the oracle symmetric benchmark of Lei et al. (2018). The rate arguments in Theorems 2–3 instead rely on the quantile regularity in Assumption A3 and the high-dimensional scaling assumptions. For unimodal but asymmetric errors, shortest valid intervals need not be symmetric, and one can form an equal-tailed split conformal interval $\widehat{C}_{\mathrm{ET}}(x) = \left[\widehat{\mu}(x)+\widehat{q}_{\alpha/2},\ \widehat{\mu}(x)+\widehat{q}_{1-\alpha/2}\right]$, using signed residuals $E_i = Y_i - \widehat{\mu}(X_i)$, where $\widehat{q}_u$ denotes the empirical u-quantile of $\{E_i\}_{i\in\mathcal{D}_{\mathrm{cal}}}$. Studying sharp efficiency guarantees for such non-symmetric targets under unimodality is an interesting direction for future work.

Finally, we return to the discussion of the validity of prediction intervals. While Theorem 1 establishes marginal validity, this guarantee is often insufficient in practice. For instance, when predicting the binding affinity

of drug candidates, marginal validity ensures coverage on average over the entire population, but it may still lead to systematic underestimation of predictive uncertainty for certain subsets of compounds. In human-centered applications, such marginally valid prediction sets can be especially untrustworthy for legally protected groups (e.g., those defined by sensitive attributes such as race, gender, or age). Therefore, stronger forms of validity are required. To this end, we leverage the definition of asymptotic conditional validity from Lei et al. (2018) to demonstrate that our proposed method possesses asymptotic conditional validity under certain conditions.

Definition 1. Denote $C_n(X_{n+1})$ as a prediction band. The prediction band has asymptotic conditional coverage at the level $(1-\alpha)$ if there exist random sets $\Lambda_n \subseteq \mathbb{R}^p$ such that $\mathbb{P}(X_{n+1} \in \Lambda_n \mid \Lambda_n) = 1 - o_p(1)$ and

$$\sup_{x_{n+1} \in \Lambda_n} \left| \mathbb{P}\left(Y_{n+1} \in C_n(X_{n+1}) \mid X_{n+1} = x_{n+1}\right) - (1-\alpha) \right| = o_p(1).$$

The following theorem shows that the prediction interval produced by our method has asymptotic conditional coverage at the level $1 - \alpha$.

Theorem 4. *Assume the linear model is correctly specified. Let $L(\cdot)$ denote the Lebesgue measure and $A \triangle B$ denote the symmetric difference between two sets. Under Assumptions A1-A3 and B1-B6, we have*

$$L(\widehat{C}_{DeCThRR}(X_{n+1}) \triangle C_o^*(X_{n+1})) = o_p(1). \tag{3.11}$$

The oracle prediction band $C_o^*(X_{n+1}))$ is conditional valid as we mentioned before. The center and width of $\widehat{C}_{\text{DeCThRR}}(X_{n+1})$ is asymptotically closed to $C_o^*(X_{n+1})$. This implies the prediction interval has asymptotic conditional coverage.

4. Numerical Experiments

We systematically compare our Debiased Conformal Threshold Ridge Regression (DeCThRR) with several competitors: the non-debiased Conformal Threshold Ridge Regression (CThRR), conformal Lasso (CLasso), conformal multi-layer perceptron (CMLP) and random forest (CRF), post-selection MLP (PMLP) and RF (PRF), and bootstrap-based intervals for threshold ridge regression (ThRR-Boot) and its debiased variant (ThRR-DeBoot) (Zhang and Politis, 2022).

4.1 Synthetic Data

To evaluate our proposed method, we generate $n = 200$ samples and split them into training ($n_{\text{train}} = 80$), calibration ($n_{\text{cal}} = 80$), and test ($n_{\text{test}} = 40$) sets, with $p \in \{500, 1000, 1500\}$. We consider four data-generating mechanisms: (i) group-sparse signals with low-rank structure (Case 1), (ii) dense weak signals concentrated in a low-dimensional subspace

(Case 2), (iii) nonlinear outcomes with polynomial interactions (Case 3), and (iv) asymptotically decaying signals under severe multicollinearity (Case 4). We also include Case 1S, a coordinate-wise sparse variant of Case 1 obtained by keeping the same design and noise but modifying β. The data-generating mechanisms and results are provided in Section S4 in the supplementary materials.

Case 1: Group Sparse Model with Low-Rank Structure. This scenario is designed to challenge methods like Lasso that assume coordinate-wise sparsity and to favor methods adept at handling correlated structures. The data is generated from $Y = \mathbf{X}\beta + \epsilon$ with $\epsilon \sim \mathcal{N}(0,\, 1.5^2\mathbf{I}_n)$. The design matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ is sampled from a zero-mean multivariate normal distribution with block-wise covariance, featuring strong intra-group correlation (0.4) and weaker inter-group correlation (0.15). The true coefficient vector is the sum of two components, i.e., $\beta = \beta_{\text{main}} + \beta_{\text{sparse}}$. The low-rank component $\beta_{\text{main}} = \mathbf{U}\theta_{\text{main}}$ is constructed by drawing an orthonormal basis $\mathbf{U} \in \mathbb{R}^{p\times r}$ with $r = 50$ from a random Gaussian matrix, and sampling $\theta_{\text{main}} \sim \mathcal{N}(\mathbf{0}, (1.2)^2\mathbf{I}_r)$. The group-sparse component β_{sparse} is supported on the first four groups, each of size 20, where coefficients in group $g = 0, 1, 2, 3$ are defined as $\beta_j^{\text{sparse}} = (1.5 - 0.3g)d_j + \eta_j$, with $\{d_j\}$ following a linear decay from 1.0 to 0.1 across the group and $\eta_j \sim \mathcal{N}(0, 0.05^2)$ providing small

perturbations, while coefficients outside the active groups are zero. This structure tests the ability of our method to leverage its Ridge component to model the low-rank part while using its thresholding component for the group-sparse part.

Case 2: Dense Weak Signal Model. The data is generated via $Y = \mathbf{X}\beta + \epsilon$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(0,\, 1.2^2\mathbf{I}_n)$. The true coefficient vector is given by $\beta = \beta_{\text{main}} + \beta_{\perp}$. The main signal component is defined as $\beta_{\text{main}} = \mathbf{U}\theta_{\text{main}}$, where $\theta_{\text{main}} \in \mathbb{R}^r$ has entries $\theta_k = 0.8\, z_k \cdot w_k$, with $z_k \sim \mathcal{N}(0,1)$ and weights w_k linearly decaying from 1 to 0.2 across coordinates. The orthonormal basis $\mathbf{U} \in \mathbb{R}^{p\times r}$ is obtained by QR decomposition of a Gaussian random matrix, and its orthogonal complement $\mathbf{U}_\perp \in \mathbb{R}^{p\times(p-r)}$ is similarly generated. The orthogonal component $\beta_\perp = \mathbf{U}_\perp\eta$, where $\eta \sim \mathcal{N}(\mathbf{0},\, 0.3^2\mathbf{I}_{p-r})$. The design matrix $\mathbf{X}$ is then constructed to align with the principal subspace by setting $\mathbf{X} = \mathbf{Z}\mathbf{U}^\top + 0.2\mathbf{E}$, where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r)^{n\times r}$, $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)^{n\times p}$, and columns of $\mathbf{X}$ are standardized to have mean zero and unit variance. This case assesses the ability to capture weak yet dense low-rank signals and to correct shrinkage bias within the principal subspace.

Case 3: Polynomial Interaction Model. This scenario evaluates robustness to model misspecification. The data-generating process is non-linear: $Y = \mathbf{X}\beta_{\text{main}} + 0.7\mathbf{X}_1\mathbf{X}_2 - 0.5\mathbf{X}_2^2 + 0.6\mathbf{X}_1\mathbf{X}_3 + \epsilon$. The linear component

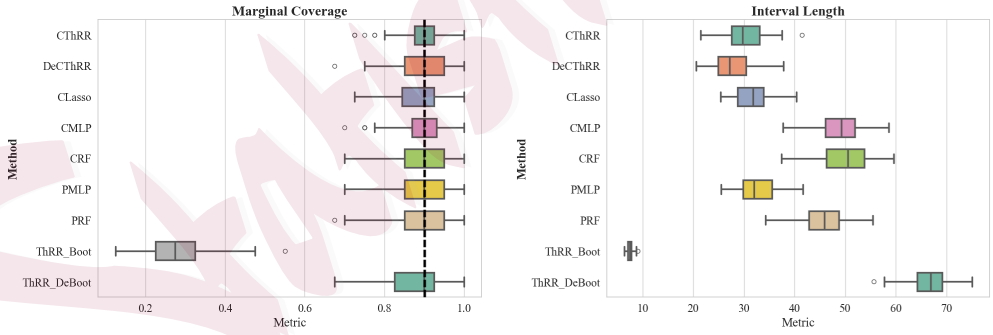
β_{main} is sparse with $s = 10$ active predictors. The design matrix $\mathbf{X}$ has a covariance structure with $\Sigma_{jk} = 0.5^{|j-k|}$. Because our base ThRR model is linear, it is misspecified for this data. This scenario tests the ability of our conformal framework, particularly with the debiased estimator, to provide robust uncertainty estimates under such misspecification.

Case 4: Asymptotic Decay Model. We generate $Y = \mathbf{X}\beta + \epsilon$, $\epsilon \sim \mathcal{N}(0, \mathbf{I}_n)$, with $\mathbf{X} \sim \mathcal{N}(0, \Sigma)$ and Toeplitz covariance $\Sigma_{jk} = 0.8^{|j-k|}$, inducing strong multicollinearity. The signal is non-sparse, with coefficients following $\beta_j = (-1)^j \exp(-j/20), \quad j = 1, \ldots, p$. This setting produces severe collinearity and gradual signal decay, leading to substantial shrinkage bias under Ridge regularization and highlighting the potential gains from debiasing.

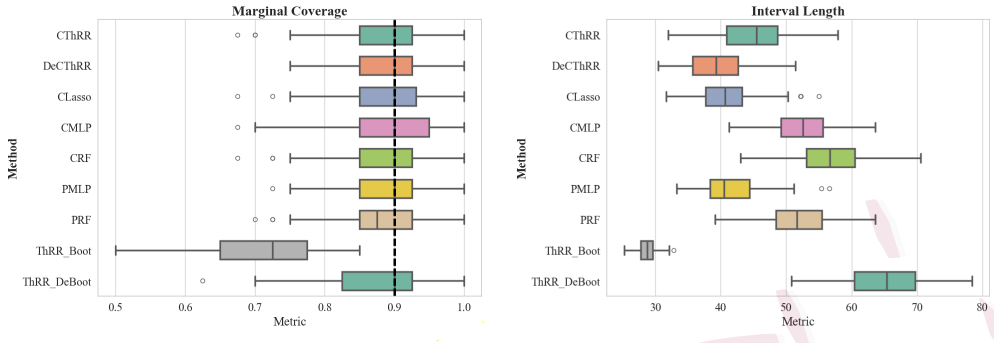
For all simulation scenarios, hyperparameters of the competing methods, for example, the regularization parameters for ThRR and Lasso, are tuned via 5-fold cross-validation on the training set. The nominal coverage level of all prediction intervals is set to $1 - \alpha = 0.90$. Each experiment is repeated over 200 independent simulations. Results for Case 1 are reported in Table 1 and Figure 2, while the results for Case 1S, Cases 2–4 and the data-generating details are provided in Section S4 of the supplementary materials.

Table 1: Average interval length, coverage (in parentheses) and time for Case 1.

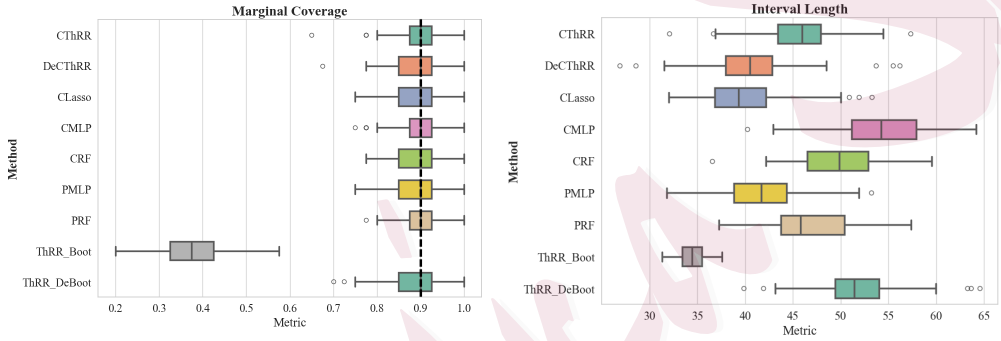
	$p = 500$	$p = 1000$	$p = 1500$	Time (s)
CThRR	30.129(0.894)	44.803(0.886)	45.802(0.897)	**0.014**
DeCThRR	**27.632(0.897)**	**39.357(0.891)**	40.651(0.881)	0.201
CLasso	31.586(0.885)	40.764(0.889)	**39.788(0.893)**	0.130
CMLP	49.075(0.892)	52.280(0.888)	54.037(0.893)	0.458
CRF	49.940(0.890)	56.508(0.884)	49.820(0.900)	1.460
PMLP	32.514(0.890)	41.388(0.893)	41.862(0.900)	0.531
PRF	45.968(0.887)	51.623(0.879)	46.940(0.899)	0.196
ThRR_Boot	7.418(0.278)	28.768(0.717)	34.450(0.776)	2.221
ThRR_DeBoot	66.777(0.880)	64.761(0.878)	51.706(0.877)	22.701



(a) Boxplot of coverage and interval width when $p = 500, n = 200$.



(b) Boxplot of coverage and interval width when $p = 1000, n = 200$.



(c) Boxplot of coverage and interval width when $p = 1500, n = 200$.

Figure 2: Average coverage and interval length of methods in Case 1.

Our main findings are summarized as follows. First and most importantly, the proposed DeCThRR method consistently improves upon its non-debiased counterpart, CThRR, across all five cases. Moreover, we analyze the performance of both methods by varying the ridge regularization parameter h_n (see Figure 3), which will be discussed in detail later. These results empirically validate our central hypothesis that explicitly correcting

for regularization bias leads to more informative prediction intervals.

In the structured sparse setting with low-rank components (Case 1) and coordinate-wise sparse (Case 1S), DeCThRR is highly competitive with the specialized CLasso method.It indicates that the stability of ridge-type estimation together with bias correction can yield more informative conformal intervals than purely sparsity-driven approaches in correlated designs. In the dense weak-signal and asymptotically decaying coefficient settings (Cases 2 and 4), which feature substantial multicollinearity, DeCThRR is designed to excel by combining the stability of ridge-type estimation with a targeted bias correction. Finally, in the model misspecification scenario (Case 3), while nonlinear methods such as CRF and PRF perform best as expected, DeCThRR remains the most competitive among the methods based on a linear working model. ThRR_Boot can severely undercover because bootstrap calibration based on the biased ThRR fit may underestimate the tail uncertainty of the prediction error. By contrast, ThRR_DeBoot partially mitigates this issue by debiasing before bootstrap calibration, typically improving coverage at the cost of wider intervals and substantially higher runtime.

In summary, while no single method dominates across all possible data structures, our proposed DeCThRR not only systematically improves upon

its baseline but also provides a robust and competitive framework for uncertainty quantification, particularly in challenging high-dimensional scenarios with correlated and non-sparse signals. We also report runtimes in the tables. CThRR is typically fastest since the solution of ThRR is closed-form. CLasso is usually next, and DeCThRR incurs only a modest overhead for debiasing. Bootstrap-based baselines are substantially more expensive due to repeated refitting.

In addition to the overall performance comparison, we conducted experiments to analyze the length of prediction intervals produced by CThRR and our proposed DeCThRR when varying the Ridge regularization parameter h_n. For this analysis, we fixed the thresholding parameter a_n at a reasonable value chosen by cross-validation and varied h_n over a wide range. Figure 3 presents the average interval length of CThRR and DeCThRR as a function of h_n for Cases 1-4 when $N = 200, p = 1500$. Additional results for $p = 500$ and $p = 1000$ are shown in the supplementary materials. The plot clearly illustrates that DeCThRR yields narrower intervals than CThRR across the entire range of $h_n > 0$. The performance gap is most significant for intermediate values of h_n, where the regularization provides necessary stability but also introduces considerable bias. This visualization directly confirms the benefit of our debiasing procedure in correcting

the bias induced by the Ridge penalty, leading to more efficient prediction intervals. The curves are noisier in Cases 3–4 because these settings are intrinsically less stable in finite samples: Case 3 is misspecified due to interaction terms, and Case 4 exhibits severe multicollinearity with a non-sparse, slowly decaying signal, both of which amplify split-to-split and Monte Carlo variability.

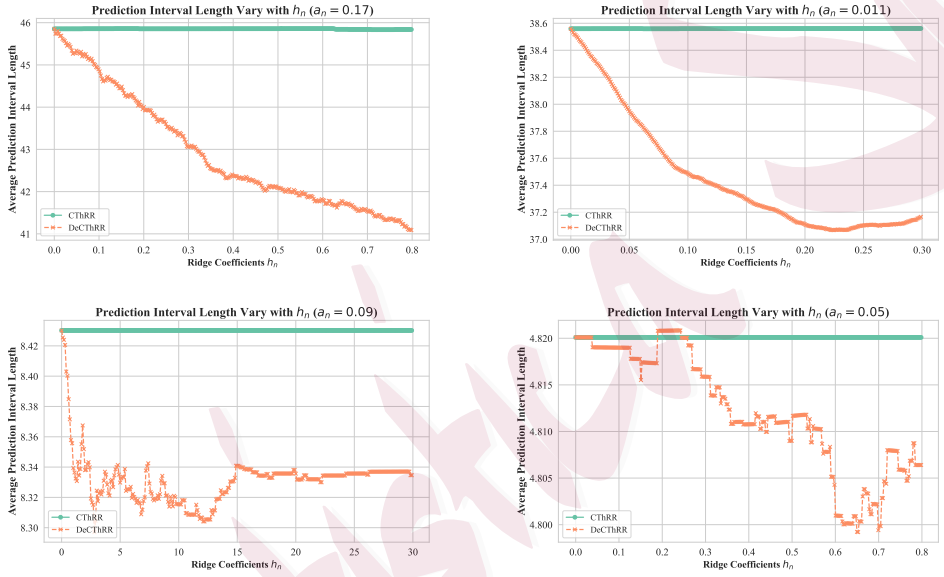


Figure 3: Prediction interval lengths of DeCThRR and CThRR across Cases 1–4 for $p = 1500$. From left to right: Case 1 (top left), Case 2 (top right), Case 3 (bottom left), and Case 4 (bottom right).

4.2 Real Data

We next evaluate the methods on the *corn* near-infrared (NIR) spectroscopy dataset. The data contain $n = 80$ corn samples and $p = 700$ spectral variables (wavelengths). The task is to predict moisture content from spectra. As is typical for NIR data, adjacent wavelengths are highly correlated, yielding strong multicollinearity. Features are standardized to zero mean and unit variance.

The experimental protocol mirrors the synthetic setup: we randomly split the data into training(40%), calibration(40%), and test sets(20%), select hyperparameters via five-fold cross-validation on the training set, and set the nominal coverage to $1 - \alpha = 0.90$. We repeat the experiment 200 times with new random splits and report empirical coverage and average interval length.

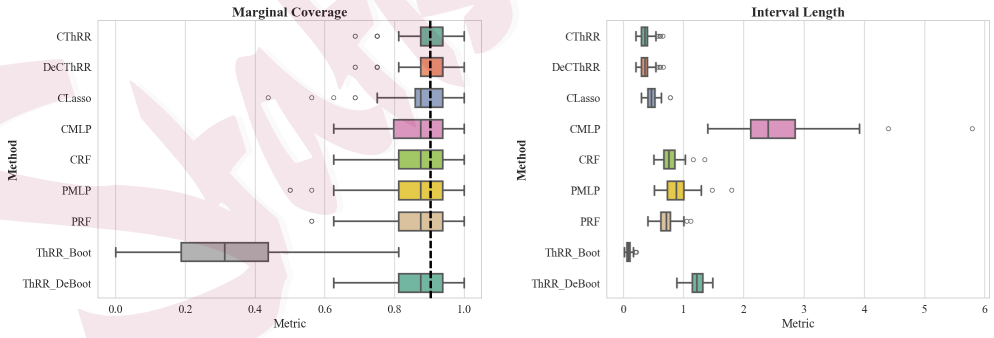


Figure 4: Average coverage and interval length of methods in corn data.

Table 2: Average coverage length, empirical interval and time for corn data.

	Avg. Coverage	Avg. Length	Avg. Time
CThRR	0.903	0.368	0.017
DeCThRR	**0.903**	**0.367**	0.045
CLasso	0.885	0.472	**0.008**
CMLP	0.853	2.532	0.111
CRF	0.871	0.774	0.475
PMLP	0.869	0.892	0.015
PRF	0.880	0.718	0.072
ThRR_Boot	0.304	0.089	0.680
ThRR_DeBoot	0.873	8.887	3.713

The results are summarized in the Table 2 and visualized in Figure 4. We first observe that the standard bootstrap method (ThRR_Boot) fails, yielding severe undercoverage (30.4%), while its debiased variant (ThRR_DeBoot) produces extremely conservative and uninformatively wide intervals. For the conformal methods, the empirical coverage rates are all close to the nominal 90% level, allowing for a fair comparison of their interval lengths. The key finding is that our proposed DeCThRR and its non-debiased counterpart CThRR are the clear best-performing methods, yielding average interval lengths of 0.367 and 0.368, respectively. They substantially out-

perform CLasso (0.472) and all other competitors by a large margin. This result strongly supports our hypothesis that the Threshold Ridge Regression framework is well-suited for data with strong multicollinearity.

In summary, our numerical experiments demonstrate the contribution of the DeCThRR framework. The debiasing step provides consistent and significant improvements in prediction interval efficiency across a variety of challenging high-dimensional settings, establishing our method as a robust and powerful tool for uncertainty quantification.

5. Conclusion

This paper introduces the Debiased Conformal Threshold Ridge Regression (DeCThRR) framework for constructing efficient prediction intervals in high-dimensional settings. By integrating a targeted debiasing step with a stable Threshold Ridge Regression estimator, our method overcomes the key trade-off in conformal prediction—avoiding the conservativeness of overfitted models while mitigating the bias of regularized estimators.

Theoretically, we establish that DeCThRR preserves the finite-sample marginal coverage guarantee of conformal prediction while achieving asymptotic efficiency and conditional validity. Extensive numerical experiments corroborate these results, showing that DeCThRR consistently yields nar-

rower prediction intervals than its non-debiased counterpart, standard conformal methods, and two-stage post-selection strategies, especially under strong multicollinearity or model misspecification. Our theory is developed under exchangeability. Since the proposed debiasing step only modifies the base predictor used in the nonconformity score, it may also be incorporated into conformal procedures for non-exchangeable settings. For example, under covariate shift one can combine DeCThRR with weighted conformal calibration via importance-weighted quantiles. We leave formal analysis of such extensions to future work.

Future research could extend this framework beyond linear models to generalized linear or nonparametric settings. Another promising direction is to jointly correct both regularization and model misspecification bias. Building upon our conditional coverage results, future work may also explore heteroscedastic extensions to attain stronger local coverage guarantees and adaptivity to varying noise levels.

Supplementary Materials

This file provides the proofs of theorems mentioned in the paper, and some additional simulation results.

Acknowledgements

This research was partially supported by National Natural Science Foundation of China (12371322, 12371265, 12401430).

References

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

Barber, R. F., E. J. Candès, A. Ramdas, and R. J. Tibshirani (2023). Conformal prediction beyond exchangeability. *The Annals of Statistics 51*(2), 816–845.

Bashari, M., A. Epstein, Y. Romano, and M. Sesia (2023). Derandomized novelty detection with FDR control via conformal e-values. In *Advances in Neural Information Processing Systems*, Volume 36, pp. 65585–65596.

Berk, R., L. Brown, A. Buja, K. Zhang, and L. Zhao (2013). Valid post-selection inference. *The Annals of Statistics 41*(2), 802–837.

Burnaev, E. and V. Vovk (2014). Efficiency of conformalized ridge regression. In *Proceedings of the 27th Conference on Learning Theory*, Volume 35, pp. 605–622. PMLR.

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2021). Distributional conformal prediction. *Proceedings of the National Academy of Sciences 118*(48), e2107794118.

Chetverikov, D. and K. Kato (2013). Gaussian approximations and multiplier bootstrap for

maxima of sums of high-dimensional random vectors. *The Annals of Statistics 41*(6), 2786–2819.

Clarté, L. and L. Zdeborová (2025). Building conformal prediction intervals with approximate message passing. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, Volume 286, pp. 798–820. PMLR.

Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap*, Volume 57 of *Monographs on Statistics and Applied Probability*. New York: Chapman and Hall/CRC.

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

Gibbs, I. and E. J. Candès (2025). Characterizing the training-conditional coverage of full conformal inference in high dimensions. *arXiv preprint arXiv:2502.20579*.

Hebiri, M. (2010). Sparse conformal predictors. *Statistics and Computing 20*, 253–266.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis* (2nd ed.). Cambridge: Cambridge University Press.

Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research 15*(1), 2869–2909.

Joshi, S., S. Kiyani, G. Pappas, E. Dobriban, and H. Hassani (2025). Likelihood-ratio regularized quantile regression: Adapting conformal prediction to high-dimensional covariate shifts. *arXiv preprint arXiv:2502.13030*.

REFERENCES

Lee, J. D., D. L. Sun, Y. Sun, and J. E. Taylor (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics 44*(3), 907–927.

Lei, J. (2019). Fast exact conformalization of the lasso using piecewise linear homotopy. *Biometrika 106*(4), 749–764.

Lei, J., M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association 113*(523), 1094–1111.

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

Liang, R., W. Zhu, and R. F. Barber (2024). Conformal prediction after efficiency-oriented model selection. *arXiv preprint arXiv:2408.07066*.

Liu, H. and B. Yu (2013). Asymptotic properties of lasso+ mls and lasso+ ridge in sparse high-dimensional linear regression. *Electronic Journal of Statistics 7*, 3124–3169.

Lu, C., A. Lemay, K. Chang, K. Höbel, and J. Kalpathy-Cramer (2022). Fair conformal predictors for applications in medical imaging. *Proceedings of the AAAI Conference on Artificial Intelligence 36*(11), 12008–12016.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics 21*(1), 255–285.

Shao, J. and X. Deng (2012). Estimation in high-dimensional linear models with deterministic design matrices. *The Annals of Statistics 40*(2), 812–831.

Tibshirani, R. J., R. F. Barber, E. J. Candès, and A. Ramdas (2019). Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems*, Volume 32, pp. 2530–2540.

Tibshirani, R. J., A. Rinaldo, R. Tibshirani, and L. Wasserman (2018). Uniform asymptotic inference and the bootstrap after model selection. *The Annals of Statistics 46*(3), 1255–1287.

Van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42*(3), 1166–1202.

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

Vovk, V., A. Gammerman, and G. Shafer (2005). *Algorithmic Learning in A Random World*. Springer.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 76*(1), 217–242.

Zhang, Y. and D. N. Politis (2022). Ridge regression revisited: Debiasing, thresholding and bootstrap. *The Annals of Statistics 50*(3), 1401–1422.

REFERENCES

Zrnic, T. and M. I. Jordan (2023). Post-selection inference via algorithmic stability. *The Annals of Statistics 51*(4), 1666–1691.

Jiamei Wu, Beijing Jiaotong University, China

E-mail: jmwu01@bjtu.edu.cn

Pan Shang, Beijing Jiaotong University, China

E-mail: 10376@bjtu.edu.cn

Yanlin Tang, East China Normal University, China

E-mail: yltang@fem.ecnu.edu.cn

Linglong Kong, University of Alberta, Canada

E-mail: lkong@ualberta.ca

Bei Jiang, University of Alberta, Canada

E-mail: bei1@ualberta.ca

Lingchen Kong, Beijing Jiaotong University, China

E-mail: lchkong@bjtu.edu.cn