Conformalized Fairness via Quantile Regression

Abstract

Fairness and impartiality have been among the predominant pursuits in human society. While rich literature on mean fairness has been established, research on quantile fairness remains sparse but vital. To fulfill great needs and advocate the significance of quantile fairness, we propose a novel framework to learn a real-valued quantile function under the fairness requirement with respect to sensitive attributes, such as race or gender, and thereby derive a reliable "fair" prediction interval. Using optimal transport and functional synchronization techniques, we establish theoretical guarantees of distribution-free coverage and exact fairness for the induced prediction interval constructed by fair quantiles. A hands-on pipeline is provided to incorporate flexible quantile regressions with an efficient fairness adjustment post-processing algorithm. We demonstrate the superior empirical performance of this approach on several benchmark datasets. Our results highlight the model's ability to uncover fairness–coverage frontiers in a variety of societal and medical applications.

1 Introduction

We are increasingly leaning on machine learning systems to tackle human problems. A primary goal is to enable intelligent algorithms to automatically make more accurate and equitable decisions, as the prevalence of unintended social bias has recently been identified as a major concern [12, 18].

There has been considerable research on algorithmic fairness focusing on conditional mean inferences [1, 9, 10, 38, 48]. However, it is equally important or perhaps even more profound to examine fairness at each specific location within a population, namely, quantile fairness, since biases and disparities may be more severe at specific quantiles, which are often overlooked. Moreover, when the data is heteroscedastic or lacks satisfactory distributional assumptions, such as Gaussianity, estimating conditional quantiles would be of greater practical value. In this paper, we propose a novel framework for searching fair quantile estimators using optimal transport techniques [2, 10] by transforming *bias-affected* distributions into an *only-fair* Wasserstein-2 barycenter through a kernel-based functional synchronization method [7, 52].

In addition, as with many high-stakes areas, such as medical science and public policy, point estimates are insufficient to inform decisions due to the potentially huge loss incurred by wrong actions. It is also crucial to provide a prediction interval that covers the true response with high probability. Toward this end, we propose a conformalized fair quantile prediction (CFQP) inspired by the works of Romano et al. [41, 42]. It achieves a finite sample, distribution-free validity, satisfies demographic parity on different quantiles, and provides an exact control of miscoverage rates, regardless of the underlying quantile algorithm used.

To make it more intuitive, we illustrate the advantages of our method with a societal instance. Some job boards require employers to provide a salary range while recruiting. Thus, an algorithm that predicts credible salary intervals based on features like job vacancies, industries, and occupational categories ... would save tremendous time and effort. Since the wage gap distribution is dramatically skewed in socially sensitive domains, for instance, those dominated by the male or the female [17, 18],

Preprint. Under review.

it is essential to consider the whole distribution using quantile regression while providing fairness constrained response ranges. Figure 1 provides a simple illustration on how to achieve fair quantiles.

The proposed method in this paper satisfies the aforementioned objectives in the example, furthermore, provides additional advantages over traditional fairness algorithms [1, 10, 51]. Firstly, it is a brand-new quantile-based framework that searches for fair quantiles based on functional synchronization and accommodates heteroscedastic data better. Furthermore, it offers a finite sample, distribution-free validity, demographic parity, and an exact control of miscoverage rates. Finally, the proposed method has full theoretical guarantees, and superiority has been demonstrated compared with current state-of-the-art methods using extensive numerical analysis.

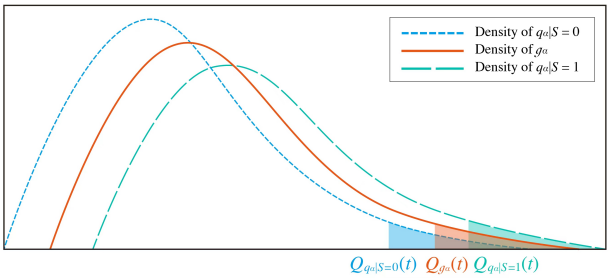

Figure 1: An illustration of quantile fairness: For a skewed and heteroscedastic quantile estimation $\{q_{\alpha,i}\}_{i=1}^N$ affacted by the sensitive attribute $S \in \{0,1\}$, for example, the lower quartile of the salary distribution, the optimal fair quantile prediction $Q_{g_\alpha}(t), t \in (0,1)$ is derived through a convex combination of the conditional quantile functions of $Q_{q_\alpha|S=0}$ and $Q_{q_\alpha|S=1}$.

Contributions and Outlines. This paper intends to bridge the gap of quantile fairness along with finite sample, distribution-free validity of conformal quantile regression. In particular,

i. We transform the problem of searching quantiles under Demographic Parity (DP) fairness to the construction of multi-marginal Wasserstein-2 barycenters via the optimal transport theory [2, 10, 19]. Furthermore, we incorporate a novel kernel smoothing step into the preceding method, which is particularly advantageous for subgroups whose sample sizes are too small to obtain reliable quantile function estimations. The post-processing algorithm is able to accommodate any off-the-shelf quantile regression.

ii. In Section 5, we propose a conformalized fair quantile regression prediction interval (CFQP) inspired by the works of Romano et al. [41, 42]. It achieves a finite sample, distribution-free validity, satisfies demographic parity on different quantiles, and provides an exact control of miscoverage rates, regardless of the underlying quantile algorithm used. The theoretical validity of prediction interval constructed by CFQP and exact DP of the fair quantile estimators are given in Section 6 and Appendix.

iii. The experimental results presented in Section 7 include a numerical comparison of the proposed CFQP and fair quantile estimation with both state-of-the-art conformal and fairness-oriented methods. By reducing the discriminatory bias dramatically, our method outperforms the state-of-the-art methods while maintaining reasonable short interval lengths.

Related works. Until recently, much fairness related work has focused on classification problems, where several quantitative measures of fairness such as demographic parity[21], equalized odds[8, 35], individual fairness[27], causality-based fairness[27, 37] have been proposed [11], leading to a variety of algorithmic classifiers that seek to neutralize the sensitive gaps.

While current approaches for building fair mean regression on a continuous quantity broadly fall into three categories. Preprocessing methods focus on transforming the data to remove any unwanted bias [5, 37, 51]. In-processing fairness constraint methods attempt to build in fairness constraints into the training step [1, 3, 27, 35]. Post-processing methods focus on modifying the trained predictor [9, 10, 34]. Few previous works concentrate on the fairness of quantiles and fair interval prediction, the most related are Yang et al. [48], where a different fairness measure rather than DP-fairness of the quantile regressor is considered, as well as Romano et al. [41], they aim to achieve equalized coverage among groups without fairness on interval endpoints.

Methodologically, integrating algorithmic fairness with Wasserstein distance based barycenter problem has been studied in [2, 9, 10, 19, 21]. Both in-processing [1, 21] and post-processing [9, 10] methods were proposed to solve classification and mean regression problems. Our work is distinct from theirs by building the DP-fairness on each population quantile and generating fairness-oriented prediction interval accordingly.

Notations. We denote by $[K]$ the set $\{1, \ldots, K\}$ for arbitrary integer K. $|\mathcal{S}|$ represents the cardinality for a finite set $\mathcal{S}$. $\mathbb{E}$ and $\mathbb{P}$ represent for the generic expectation and probability and $\mathbb{1}\{\cdot\}$ is the indicator function. Let $Z_n, n = 1, 2, \ldots,$ be a sequence of random variables, and $\{k_n\}_{n=1}^{\infty}$ be a sequence of positive numbers, we say that $Z_n = O_p(k_n)$, if $\lim_{T\to\infty} \limsup_{n\to\infty} \mathbb{P}(|Z_n| > Tk_n) = 0$, then $Z_n/k_n = O_p(1)$. For two random variables A and B, we write $A \overset{d}{=} B$ to denote their equality in distribution.

2 Problem statement

Consider the regression problem where a "sensitive characteristic" S is available, which by the U.S. law [19, 41] can be enumerated as sex, race, age, disability, etc. One observes the triplets $(X_1, S_1, Y_1), \ldots, (X_n, S_n, Y_n)$, where each $Z_i = (X_i, S_i, Y_i), i = 1, \ldots, n$ is a random variable in $\mathbb{R}^p \times [K] \times \mathbb{R}$. Let $\mathbb{P}$ be the joint distribution of Z, we assume that all the samples $\{Z_i\}_{i=1}^{n+1}$ are drawn exchangeably. For a more specific case, they may be drawn i.i.d. from $\mathbb{P}$.

We aim to construct a marginal distribution-free prediction band $C(X_{n+1}, S_{n+1}) \subseteq \mathbb{R}$ that is likely to cover the unknown response Y_{n+1} with finite-sample (nonasymptotic) validity. Formally, given a desired miscoverage rate α, the predicted interval satisfies

$$\mathbb{P}\{Y_{n+1} \in C(X_{n+1}, S_{n+1})\} \geq 1 - \alpha \tag{1}$$

for any joint distribution $\mathbb{P}$ and any sample size n, while the left and right endpoint of $C(X_{n+1}, S_{n+1})$ satisfies the fairness constraint of Demographic Parity concerning the sensitive variable S.

3 Quantile Regression and Conformal Prediction

We here recall the CQR approach for finite sample, distribution-free prediction interval inference. Quantile regression was proposed by Koenker and Bassett [25] to estimate the α-th quantile of the conditional distribution of Y given $\tilde{X} := (X, S)$ for some quantile level $\alpha \in (0, 1)$, since then it becomes more and more pervasive with various applications, such as providing prediction intervals, detecting outliers, or perceiving the entire distribution [33]. As a brief introduction to the quantile regression. Denote the conditional cumulative distribution of Y given $\tilde{X}$ by $F(y \mid \tilde{X} = \tilde{x}) := \mathbb{P}\{Y \leq y \mid \tilde{X} = \tilde{x}\}$. The α-th conditional quantile prediction is defined as

$$q_\alpha(\tilde{x}) := \inf\{y \in \mathbb{R} : F(y \mid \tilde{X} = \tilde{x}) \geq \alpha\}.$$

Quantile regression can be cast as an optimization problem[33], by minimizing the expected check loss function $E(\rho_\alpha) = E[\rho_\alpha(y, q)|\tilde{X} = \tilde{x}]$, where

$$\rho_\alpha(y, q) = \begin{cases} \alpha|y - q| & \text{if} \quad q > 0, \\ (1 - \alpha)|y - q| & \text{if} \quad q \leq 0. \end{cases} \tag{2}$$

Both parametric and nonparametric approaches have been developed to compute the conditional quantile predicitons $q_\alpha(\tilde{x})$ efficiently due to the convex nature of the optimization problem [33, 36, 39]. Common non-parametric approaches are among quantile Smoothing Splines[20, 26, 33], quantile gradient boosting[49], random forests[33], as well as quantile neural networks[44]. Under regularity conditions and for certain models, estimates of conditional quantile functions via the check loss or other alternatives are renowned to be asymptotically consistent [25, 33, 43, 53].

Quantile regression thus offers a principled way of judging the reliability of predictions by building a prediction interval for the new observation $(\tilde{X}_{n+1}, Y_{n+1})$. In contrast to asymptopia, Romano et al. [42, 41] brought forward the conformalized quantile regression (CQR) by combining the merits of robust quantile regression with conformal prediction; thus finite sample validity in (1) is guaranteed. Inspired by the split conformal method, a split CQR likewise starts with splitting the data into a

proper training set and a calibration set, indexed by $\mathcal{I}_1, \mathcal{I}_2$ respectively. Given any quantile regression algorithm $\mathcal{Q}$, we then fit two conditional quantile functions $\hat{f}_{\alpha_{\rm lo}}$ and $\hat{f}_{\alpha_{\rm hi}}$ on the proper training set:

$$\{\hat{q}_{\alpha_{\rm lo}}, \hat{q}_{\alpha_{\rm hi}}\} \leftarrow \mathcal{Q}\left(\left\{\left(\tilde{X}_i, Y_i\right) : i \in \mathcal{I}_1\right\}\right)$$

Then the conformity scores are calculated to quantify the error made by the plug-in prediction interval $\hat{C}(\tilde{x}) = [\hat{q}_{\alpha_{\rm lo}}(\tilde{x}), \hat{q}_{\alpha_{\rm hi}}(\tilde{x})]$. We evaluate the scores on the calibration set as

$$E_k := \max\left\{\hat{q}_{\alpha_{\rm lo}}(\tilde{X}_k) - Y_k, Y_k - \hat{q}_{\alpha_{\rm hi}}(\tilde{X}_k)\right\}$$

for each $k \in \mathcal{I}_2$, where both undercoverage and overcoverage of the interval are taken into consideration [42]. Given a new input data $\tilde{X}_{n+1}$, we construct the prediction interval for Y_{n+1} as

$$C\left(\tilde{X}_{n+1}\right) = \left[\hat{q}_{\alpha_{\rm lo}}\left(\tilde{X}_{n+1}\right) - Q_{1-\alpha}\left(E, \mathcal{I}_2\right), \hat{q}_{\alpha_{\rm hi}}\left(\tilde{X}_{n+1}\right) + Q_{1-\alpha}\left(E, \mathcal{I}_2\right)\right]$$

where $Q_{1-\alpha}(E, \mathcal{I}_2) := (1-\alpha)(1 + 1/|\mathcal{I}_2|)$-th empirical quantile of $\{E_k : k \in \mathcal{I}_2\}$ conformalizes the plug-in prediction interval. Note that the constructed interval $C(\tilde{X}_{n+1})$ could be highly influenced by the sensitive variable S.

4 Fairness risk measures

In this section, we introduce the quantitative definition of DP on fair regression and connect the DP-fairness with quantile regressor q_α. We invoke the result that q_α can be projected to fairness using optimal transport.

Given a fixed quantile level α (it may refer to $q_{\alpha_{\rm lo}}$ or $q_{\alpha_{\rm hi}}$ in our case), let $q_\alpha : \mathbb{R}^p \times [K] \to \mathbb{R}$ represent an arbitrary conditional quantile predictor. Denote by $\nu_{q_\alpha|s}$ the distribution of $(q_\alpha(X, S) \mid S = s)$, the Cumulative Distribution Function (CDF) of $\nu_{q_\alpha|s}$ is given by

$$F_{\nu_{q_\alpha|s}}(t) = \mathbb{P}(q_\alpha(X, S) \leq t \mid S = s). \tag{3}$$

The quantile function $Q_{\nu_{q_\alpha|s}} = F^{-1}_{\nu_{q_\alpha|s}} : [0, 1] \to \mathbb{R}$,namely, the generalized inverse of $F_{\nu_{q_\alpha|s}}$, can thus be defined as for all levels $t \in (0, 1]$,

$$Q_{\nu_{q_\alpha|s}}(t) = \inf\{y \in \mathbb{R} : F_{\nu_{q_\alpha|s}}(y) \geq t\} \text{ with } Q_{\nu_{q_\alpha|s}}(0) = Q_{\nu_{q_\alpha|s}}(0+). \tag{4}$$

To simplify the notations, we will write $F_{q_\alpha|s}$ and $Q_{q_\alpha|s}$ instead of $F_{\nu_{q_\alpha|s}}$ and $Q_{\nu_{q_\alpha|s}}$ respectively, for any prediction rule q_α.

In the following, we introduce the definition of Demographic Parity (DP), which is most commonly used in the context of fairness research [1, 9, 10, 35]. This concept could be thought of as a stronger version of the US Equal Employment Opportunity Commission's "four-fifths rule," which says "selection rate for any race, sex, or ethnic group must be at least $4/5$ or 80% of the rate for the group with the highest rate,"[1], and in turn, reflects the reason for pursuing DP in quantile prediction.

Definition 1 (Demographic Parity)**.** An arbitrary prediction $g : \mathbb{R}^d \times [K] \to \mathbb{R}$ satisfies demographic parity under a distribution $\mathbb{P}$ over (X, S, Y) if $g(X, S)$ is statistically independent of the sensitive attribute S, namely, for every $s, s' \in [K]$,

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(g(X, S) \leq t \mid S = s) - \mathbb{P}\left(g(X, S) \leq t \mid S = s'\right)| = 0.$$

Demographic Parity (DP) requires the predictions to be independent of the sensitive attribute, and it demands the Kolmogorov-Smirnov distance [28] (the difference between CDFs measured in the l_∞ norm) between $\nu_{g|s}$ and $\nu_{g|s'}$ to vanish for all categories s, s'. Proposition 1 pronounces that the optimal fair quantile predictor for an input (x, s) is obtained by a nonlinear transformation of the vector $[q_\alpha(x, s)]_{s=1}^K$ linking to a Wasserstein barycenter problem[2, 10].

Definition 2 (Wasserstein-2 distance)**.** Let μ and ν be two univariate probability measures with finite second moments. The squared Wasserstein-2 distance between μ and ν is defined as

$$\mathcal{W}_2^2(\mu, \nu) = \inf\left\{\int_{\mathbb{R}\times\mathbb{R}} |x - y|^2 d\gamma(x, y), \gamma \in \Gamma_{\mu,\nu}\right\}$$

where $\Gamma_{\mu,\nu}$ is the set of probability measures (couplings) on $\mathbb{R} \times \mathbb{R}$ having μ and ν as marginals.

Proposition 1 (Fair optimal prediction [10])**.** Assume, for each $s \in [K]$, that the univariate measure $\nu_{q_\alpha|s}$ has a density and let $p_s = \mathbb{P}(S = s)$. Then,

$$\min_{g_\alpha \text{ is fair}} \mathbb{E}\left(q_\alpha(X,S) - g_\alpha(X,S)\right)^2 = \min_{\nu} \sum_{s\in[K]} p_s \mathcal{W}_2^2\left(\nu_{q_\alpha|s}, \nu\right). \tag{5}$$

Moreover, if g_α and ν solve the l.h.s. and the r.h.s. problems respectively, then $\nu = \nu_{g_\alpha}$ and specifically,

$$g_\alpha(x,s) = \sum_{s'\in[K]} p_{s'} Q_{q_\alpha|s'} \circ F_{q_\alpha|s} \circ q_\alpha(x,s). \tag{6}$$

The explicit closed form solution comes from [2, 10, 16], which relies on the classical characterization of optimal coupling in one dimension of the Wasserstein-2 distance. A rigorous proof is given in [10, 21]. It shows that a minimizer g_α of the L_2-risk can be used to construct ν and vice-versa, given ν, there is a explicit expression Eq. (6) for the multi-marginal Wasserstein barycenter [2].

5 Conformal fair quantile prediction (CFQP)

We formally describe our proposed conformal fair prediction (CFQP) framework for constructing DP fairness constrained prediction intervals in this part. A kernel smoothing quantile function is introduced during the functional synchronization, which can improve the estimation when some subgroups are too small to give reliable sample quantile function estimations.

After separating the data into a proper training set with indices $\mathcal{I}_1$ and a calibration set $\mathcal{I}_2$, we start with fitting an arbitrary quantile regression algorithm $\mathcal{Q}$ on $\mathcal{I}_1$,

$$\{\hat{q}_{\alpha_{\text{lo}}}, \hat{q}_{\alpha_{\text{hi}}}\} \leftarrow \mathcal{Q}\left(\left\{\left(\tilde{X}_i, Y_i\right) : i \in \mathcal{I}_1\right\}\right).$$

In the next step, before conducting the conformalization, it is essential to remove the effect of sensitive variable for the predicted $\{\hat{q}_{\alpha_{\text{lo}}}(\tilde{X}_i), \hat{q}_{\alpha_{\text{hi}}}(\tilde{X}_i)\}, \forall i \in \mathcal{I}_2$. By Proposition 1, the optimal fair quantiles take the form of Eq. (6). Therefore, we propose an empirical optimal fair quantile estimator $\hat{g}_\alpha$ that relies on the plug-in principle. In particular, Eq.(6) indicates that for each quantile level α and each category $s \in [K]$, we only need estimators for the regression function $\hat{q}_\alpha$, the proportion $\hat{p}_s$, the cumulative distribution function $F_{\hat{q}_\alpha|s}$ and the quantile function $Q_{\hat{q}_\alpha|s}$.

Note that we can empirically estimate the CDF and quantile function for each sensitive group in the calibration set $\mathcal{I}_2$ separately. Hence for each quantile level α, let $N_s := |\mathcal{I}_2^s|$, and the quantile estimators $(\hat{q}_1^s, \hat{q}_2^s, \ldots, \hat{q}_{N_s}^s)$[1] are calculated through the fitted quantile regression $\mathcal{Q}$ with training set $\mathcal{I}_1$. We define the augmented random variable for each data point in $\mathcal{I}_2$,

$$\tilde{q}_i^s := \hat{q}_i^s + U_i^s([-\sigma, \sigma]) \quad \forall i \in \mathcal{I}_2^s, s \in [K],$$

where U_i^s are i.i.d. random variables, uniformly distributed on $[-\sigma, \sigma]$ for some small positive σ, and independent from all the previously introduced random variables. it serves as a smoothing random variable, for the random variables $\tilde{q}_i^s := \hat{q}_i^s + U_i^s$ are i.i.d. continuous for any $\mathbb{P}(Y|\tilde{X})$ and $\mathcal{Q}$. Otherwise, the original $\hat{q}_i^s$ might have atoms resulting in a non-zero probability to observe ties in the group $\{q_i^s\}$ for $s = 1, \ldots, K$. This trick, also called jittering [6, 10] is often used for data visualization for tie-breaking. Using the above quantities, we build the CDF and quantile function estimators for each subgroup $s' \in [K]$ as follows,

$$\hat{F}_{q_\alpha|s'}(t) = N_s^{-1} \sum_{i=1}^{N_s} \mathbb{1}\left\{\tilde{q}_i^{s'} \le t\right\}, \quad \text{for all } t \in \mathbb{R}, \tag{7}$$

$$\hat{Q}_{2,q_\alpha|s'}(t) = \int_0^1 \hat{F}^{-1}_{q_\alpha|s'}(v) K_h(t-v)dv, \quad t \in (0,1). \tag{8}$$

The smoothed kernel estimator Eq.(8) was firstly proposed by Cheng and Parzen [7], where $K_h(\cdot) = K(\cdot/h)/h$ is a kernel function chosen as a probability density function that is symmetric around zero with bandwidth parameter $h > 0$.

[1] $\hat{q}_i^s$ depends on the quantile level α, we suppress α for notational simplification.

If the quantile functions $Q_{2,q_\alpha|s'}$ is differentiable, the derivative $Q'_{s'}(t) := Q'_{2,q_\alpha|s'}(t)$ for $t \in (0,1)$ is the quantile density function [7, 52]. We hereby give an estimation bound for Eq. (8) using kernel smoothing. For this purpose, we invoke the following conditions that can also be found in Zhang and Müller [52]:

Condition (A1). The quantile density functions $Q'_{s'}, s' = 1, \ldots, K$, are twice continuously differentiable in $(0,1)$, and satisfy $\inf_{t\in[0,1]} Q'_{s'}(t) \geq c_0 > 0$, for a constant $c_0 > 0, \forall s'$. There is a $\gamma > 0$ such that $\sup_{u\in[0,1]} u(1-u)|J_{s'}(u)| \leq \gamma$ for all s', where $J_{s'}(u) := [\mathrm{d}\log Q'_{s'}(u)/\mathrm{d}u]$.

Condition (A2). There exists $0 < L_0 < \infty$, such that $\sup_{u\in[0,1]} |\int_0^1 Q'_{s'}(t)K_h(u-t)\mathrm{d}t| \leq L_0, \forall k$.

Condition (A3). The kernel functions K is probability density functions which is symmetric around 0. For any function f that is at least twice continuously differentiable in $(0,1)$, it holds that for a $\rho > 1/2$, $\limsup_{N\to\infty} h^2 N^\rho < \infty$, and $\sup_{u\in[a,b]\subset(0,1)} \left| f(u) - \int_0^1 f(t)K_h(u-t)\mathrm{d}t \right| = O\left(N^{-\rho}\right)$, where N is defined to ensure that the numbers of measurement asymptotically increases in the same way across the groups, we assume that there exists a sequence $N = N(n)$ with $N \to \infty$, as $n \to \infty$, such that ${}^{N_{s'}}/_{N} \to \tau_{s'}$ for positive constants, and $0 < c_0 \leq \inf_{1\leq s'\leq K} \tau_{s'} \leq \sup_{1\leq s'\leq K} \tau_{s'} < C_0 < \infty, s' = 1, \ldots, K$.

Proposition 2. Under conditions (A1), (A2), and (A3), we have

$$\sup_{s'} \sup_{t\in[0,1]} \left| \hat{Q}_{2,q_\alpha|s'}(t) - Q_{q_\alpha|s'}(t) \right| = O_p\left(N^{-1/2}\right), \quad s' = 1, \ldots, K.$$

The motivation for including a smoothing step is twofold: First, smoothing the quantile function eliminates the troublesomeness in defining arbitrary quantiles from the empirical one when the sample sizes of subgroups are small. Second, the proposed kernel smoothing improves second-order efficiency by alleviating the so-called relative deficiency [14, 52].

Remark 1. One can utilize various kernels such as the Gaussian or Epanechnikov kernel with adaptive bandwidth for better practical performance. Furthermore, other smoothing methods such as splines or local linear fitting can be applied with equal effectiveness.

Consequently, for each quantile level α, the functional synchronized quantile estimator is

$$\hat{g}_{\alpha,i} = \sum_{s'=1}^{K} \hat{p}_{s'} \hat{Q}_{2,q_\alpha|s'} \circ \hat{F}_{q_\alpha|s} \circ \tilde{q}_i^s, \quad \forall i \in \mathcal{I}_2. \tag{9}$$

The proposed estimator can be deemed as the empirical counterpart with additional randomization of the explicit fair optimal formula Eq.(6).

To conformalize the adjusted fair quantiles Eq (9), we need to compute the conformity scores E_i for each $i \in \mathcal{I}_2$ that quantify the error made by the plug-in fair prediction interval $\hat{C}^g(\tilde{x}) = [\hat{g}_{\alpha_{\mathrm{lo}}}(\tilde{x}), \hat{g}_{\alpha_{\mathrm{hi}}}(\tilde{x})]$. The scores are evaluated on the calibration set as

$$E_i := \max\{\hat{g}_{\alpha_{\mathrm{lo}},i} - Y_i, Y_i - \hat{g}_{\alpha_{\mathrm{hi}},i}\}. \tag{10}$$

At the last stage, for a new data point $\tilde{X}_{n+1} = (x, s)$, and $\alpha \in \{\alpha_{\mathrm{lo}}, \alpha_{\mathrm{hi}}\}$, by defining

$$\tilde{q}_{1,i}^s = \hat{q}_i^s + U_i^s([-\sigma, \sigma]) \quad \forall i \in \mathcal{I}_1^s, \text{ and}$$
$$\tilde{q}_\alpha(x,s) = \hat{q}_\alpha(x,s) + U([-\sigma,\sigma]),$$

we use the empirical CDF of training set [2]

$$\hat{F}_{1,q_\alpha|s}(t) := \frac{1}{|\mathcal{I}_1^s|+1} \left(\sum_{i=1}^{|\mathcal{I}_1^s|} \mathbb{1}\left\{\tilde{q}_{1,i}^s < t\right\} + U([0,1]) \left(1 + \sum_{i=1}^{|\mathcal{I}_1^s|} \mathbb{1}\left\{\tilde{q}_{1,i}^s = t\right\}\right)\right) \tag{11}$$

to estimate the location $\hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x,s)$. Thus the fair quantile estimator is built as follows

$$\hat{g}_\alpha(x,s) = \sum_{s'=1}^{K} \hat{p}_{s'} \hat{Q}_{2,q_\alpha|s'} \circ \hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x,s), \forall \alpha \in \{\alpha_{\mathrm{lo}}, \alpha_{\mathrm{hi}}\}. \tag{12}$$

[2] Still, $\hat{q}_i^s$ depends on quantile level α.

The fair prediction interval for Y_{n+1} is constructed as

$$C(\tilde{X}_{n+1}) = [\hat{g}_{\alpha_{\rm lo}}(x,s) - Q_{1-\alpha}(E,\mathcal{I}_2), \hat{g}_{\alpha_{\rm hi}}(x,s) + Q_{1-\alpha}(E,\mathcal{I}_2)], \tag{13}$$

where $Q_{1-\alpha}(E,\mathcal{I}_2) := (1-\alpha)(1+1/|\mathcal{I}_2|)$-th empirical quantile of $\{E_i : i \in \mathcal{I}_2\}$ will adjust the plug-in fair prediction interval. We present the pseudo-codes of CFQP as well as the construction of $\hat{g}_\alpha$ for Eq. 9 in Algorithm 1, 2 respectively.

Algorithm 1 Split Conformal Fair Prediction (CFQP)

Input: $\mathcal{D} = \{(X_i, S_i, Y_i)\}_{i=1}^n$; miscoverage level $\alpha \in (0,1)$; quantile regression algorithm $\mathcal{Q}$.
1: Randomly split $[n]$ into disjoint proper training and calibration indices $\mathcal{I}_1, \mathcal{I}_2$.
2: Fit two conditional quantile functions on the training set $\{\hat{q}_{\alpha_{\rm lo}}, \hat{q}_{\alpha_{\rm hi}}\} \leftarrow \mathcal{Q}(\{(X_i, S_i, Y_i), i \in \mathcal{I}_1\})$.
3: Call functional Synchronization (Algorithm 2) to calculate $\{\hat{g}_{\alpha_{\rm lo}}, \hat{g}_{\alpha_{\rm hi}}\}$ for each $i \in \mathcal{I}_2$.
4: Compute $E_i \leftarrow \max\{\hat{g}_{\alpha_{\rm lo}}(X_i) - Y_i, Y_i - \hat{g}_{\alpha_{\rm hi}}(X_i)\}$ for $\forall\, i \in \mathcal{I}_2$.
5: Compute $Q_{1-\alpha}(E,\mathcal{I}_2) \leftarrow (1-\alpha)(1+{}^1\!/_{|\mathcal{I}_2|})$-th empirical quantile of $\{E_i : i \in \mathcal{I}_2\}$.
6: For a new test point (x,s), compute $\{\hat{g}_{\alpha_{\rm lo}}(x,s), \hat{g}_{\alpha_{\rm hi}}(x,s)\}$ through Algorithm 2
Output: Fair prediction interval $C(x,s) = [\hat{g}_{\alpha_{\rm lo}}(x,s) - Q_{1-\alpha}(E,\mathcal{I}_2), \hat{g}_{\alpha_{\rm hi}}(x,s) + Q_{1-\alpha}(E,\mathcal{I}_2)]$ for $(X_{n+1}, S_{n+1}) = (x,s)$.

Algorithm 2 Functional Synchronization

Input: Calibration set $\{(X_i, S_i)\}_{i\in\mathcal{I}_2}$ or new point (x,s); base quantile estimator $\mathcal{Q}$; slack parameter σ; training set $\{(X_i, S_i)\}_{i\in\mathcal{I}_1}$;
1: **if** Calibration set $\{(X_i, S_i)\}_{i\in\mathcal{I}_2}$ **then**
2: **for** $\alpha \in \{\alpha_{\rm lo}, \alpha_{\rm hi}\}$ **do**
3: $\{\tilde{q}_\alpha(X_i, S_i)\} \leftarrow \{q_\alpha(X_i, S_i) + U([-\sigma,\sigma])\}_{i\in\mathcal{I}_2}$ ▷ $U([-\sigma,\sigma])$ are used for tie-breaking
4: **for** $s' \in [K]$ **do**
5: Compute $\hat{F}_{q_\alpha|s'}(t)$, and $\hat{F}^{-1}_{2,q_\alpha|s'}(t)$ by Eq. (7) and (8).
6: Obtain $\hat{g}_\alpha(X_i, S_i) \leftarrow \sum_{s'=1}^K \hat{p}_{s'} \hat{F}^{-1}_{2,q_\alpha|s'} \circ \hat{F}_{q_\alpha|s'} \circ \tilde{q}_\alpha(X_i, S_i),\ \forall i \in \mathcal{I}_2$
7: **end for**
8: **end for**
9: **else if** New test point (x,s) **then**
10: **for** $\alpha \in \{\alpha_{\rm lo}, \alpha_{\rm hi}\}$ **do**
11: $\{\tilde{q}^s_{1,\alpha}\} \leftarrow \{\hat{q}^s_\alpha + U([-\sigma,\sigma])\}_{i\in\mathcal{I}^s_1}$ and $\tilde{q}_\alpha(x,s) \leftarrow q_\alpha(x,s) + U([-\sigma,\sigma])$
12: Compute $\hat{g}_\alpha(x,s) \leftarrow \sum_{s'=1}^K \hat{p}_{s'} \hat{F}^{-1}_{2,q_\alpha|s'} \circ \hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x,s)$ by Eq. (8) and (7)
13: **end for**
14: **end if**
Output: fair quantile prediction $\hat{g}_\alpha$ for calibration set or new test point (x,s).

6 Theoretical results

Theorem 1. If $(\tilde{X}_i, Y_i), i = 1, \ldots, n+1$ are exchangeable, then the prediction interval $C(\tilde{X}_{n+1})$ constructed by the split CFQP algorithm satisfies

$$\mathbb{P}\{Y_{n+1} \in C(\tilde{X}_{n+1})\} \geq 1 - \alpha.$$

Moreover, if the conformity scores E_i are almost surely distinct, the prediction interval is nearly exactly calibrated,

$$\mathbb{P}\{Y_{n+1} \in C(\tilde{X}_{n+1})\} \leq 1 - \alpha + 1/(|\mathcal{I}_2| + 1).$$

Remark 2. We give an extension for the conformalization step which allows coverage errors to be spread arbitrarily over the left and right tails as Corollary 1 in Appendix. Controlling the left and right tails independently yields a stronger coverage guarantee.

Theorem 2 (Demographic parity guarantee)**.** For any joint distribution $\mathbb{P}$ of (X, S, Y), any $\sigma > 0$, as well as the base quantile estimator $\hat{q}_\alpha : \mathbb{R}^p \times [K] \to \mathbb{R}$ constructed on labeled data, the estimator $\hat{g}_\alpha$ defined in Eq. (12) satisfies

$$(\hat{g}_\alpha(X,S) \mid S = s) \stackrel{d}{=} (\hat{g}_\alpha(X,S) \mid S = s') \quad \forall s, s' \in [K].$$

To prove Theorem 2, we leverage distribution-free properties on rank and order statistics presented in Lemma 3in the supplementary material. Theorem 2 and the estimator $\hat{g}$ improve upon the estimator of Chzhen et al. [10], Chzhen and Schreuder [9], for which only approximate fairness is established.

7 Experiments

To evaluate our proposed method[3], we report the performance of post-processing fairness adjustment on quantiles through the following four benchmark datasets:

Law School (LAW)[47]: The dataset contains 20,649 examples aiming to predict students' GPA based on their information and capacities, with gender as the sensitive attribute (male vs. female).
Community&Crime (CRIME)[40]: This dataset contains socio-economic, law enforcement, and crime data about communities in the US with 1994 examples. The task is to predict the number of violent crimes per 100,000 population with race as the sensitive attribute.
MEPS 2016 (MEPS)[15, 42]: The Medical Expenditure Panel Survey 2016 dataset contains information on individuals and their utilization of medical services. The goal is to predict the health care system utilization score of each individual by their features. There are 15656 examples on $p = 41$ features with race as the sensitive attribute (nonwhite vs. white).
Government Salary (GOV)[37, 38]: The government salary dataset is collected from the 2018 American Community Survey by the US Census Bureau. The yearly salary for over 200 hundred examples is the response variable, and employee race (7 categories) is identified as the sensitive attribute.

We measure the violation of DP-fairness of the quantiles required by Definition 1 through the empirical Kolmogorov-Smirnov (KS) distance. The value represents the disparity between groups $\mathcal{Z}^s = \{(X, S, Y) \in \mathcal{Z} : S = s\}$ for all $s \in [K]$,

$$\mathrm{KS}(g_\alpha) = \max_{s,s' \in [K]} \sup_{t \in \mathbb{R}} \left| \frac{1}{|\mathcal{Z}^s|} \sum_{(X,S,Y) \in \mathcal{Z}^s} \mathbb{1}\{g_\alpha(X,S) \leq t\} - \frac{1}{|\mathcal{Z}^{s'}|} \sum_{(X,S,Y) \in \mathcal{Z}^{s'}} \mathbb{1}\{g_\alpha(X,S) \leq t\} \right|.$$

Experiment results: In Table 1, we report the average performance of the proposed CFQP over 200 randomly training-test splits as well as the baseline model CQR by the coverage rate, length of prediction interval, and the KS distance of the interval endpoint. 80% of the examples are used for training, and the remaining 20% to test. We split the training part into proper training and calibration sets of equal sizes. Throughout the experiments, the nominal miscoverage rate is fixed to $\alpha = 0.1$. Among preexisting quantile algorithms, we select three leading variants: **linear model**[25], **random forests** [33] and **neural networks** [44]. Overall, our CFQP likewise CQR constructs prediction bands attaining desirable coverage around 90%, as claimed in Theorem 1. Random forest based approaches tend to be slightly more conservative than the other two w.r.t the coverage rate among all four datasets.

In the KS column concerning the DP-fairness of interval endpoints, our CFQP method greatly reduces the discriminatory bias (quantified by KS) by 70% up to 90% compared to that of CQR. In addition, the lengths of the prediction intervals mostly remain the same except for the Crime dataset, which is due to the reason of its inherent high discriminatory bias among sensitive groups.

Figure 2 presents the comparison of our post-processing fairness adjustment procedure on quantiles $\hat{g}_\alpha$ on the test set $\mathcal{Z}_{test} = \{(X_i, S_i, Y_i)\}_{i=1}^{n_{test}}$ with some state-of-the-art fairness algorithms. Since most of the algorithms are targeted at mean prediction, there is no direct comparison with our quantile fairness method; we accordingly modified the existing methods into quantile versions for comparison. The points in Figure 2 represents the mean of 200 repeated experiments with x-axis as KS distance and y-axis as Mean Absolute Error(MAE), defined as $\mathrm{MAE}(g_\alpha) = {}^1\!/_{n_{test}} \sum_{\mathcal{Z}_{test}} |Y - g_\alpha(X, S)|$ to measure the prediction error of quantiles, the bars is the standard error on both axis. The optimal point should be located at the bottom left corner of the graph, where the smaller KS distance as well as smaller MAE are achieved. In each subplot, our method consistently performs better with the smallest KS distance while keeping the MAE equal or slightly higher than the others or unfair version. Overall, CFQP is more robust according to the standard errors over experiment repetitions. Note that due to the highly right skewness of real datasets, the MAE of the upper quantile estimation is larger than that of the lower quantile for all approaches as well as datasets.

8 Conclusion and future work

Conformal fair quantile regression is a novel approach for creating fair prediction intervals that attain valid coverage and reach independence with sensitive attributes while making minimal modifica-

[3] we utilize the local linear fitting smoothing method in the experiments.

tions to the quantile endpoints simultaneously. It becomes superior within heteroskedastic and/or asymmetric datasets and robust to outliers.

CFQP is supported by rigorous distribution-free coverage and exact DP-fairness guarantees, as proved in theoretical parts. We conducted several real data examples demonstrating the effectiveness of our method in achieving exact coverage, while imposing DP-fairness in practice. The method outperforms several state-of-the-art approaches by comparison.

A limitation in our numerical experiments is that we simply utilize the local linear smoothing method in defining quantile functions of the subgroups; we believe incorporating flexible kernel smoothing approaches [52] would improve the experimental performances. In the future, it would be valuable to extend our methodology to the case when we are not allowed to use the sensitive feature via counterfactual causal models [37].

	LAW				CRIME			
	Coverage	Length	KS(lo)	KS(hi)	Coverage	Length	KS(lo)	KS(hi)
Ln-CQR	90.16±0.47	0.46±.004	0.39±0.03	0.11±0.02	90.22±1.88	1.30±0.05	0.62±0.06	0.53±0.06
Ln-CFQP	90.02±0.51	0.46±.004	0.02±0.01	0.02±0.01	90.44±1.84	1.64±0.05	0.11±0.03	0.12±0.04
RF-CQR	90.25±0.55	0.39±.005	0.20±0.02	0.15±0.02	90.27±1.66	1.15±0.03	0.64±0.05	0.59±0.05
RF-CFQP	90.11±0.48	0.38±.004	0.02±.008	0.02±.009	90.34±1.84	1.54±0.04	0.12±0.04	0.12±0.03
NN-CQR	90.00±0.50	0.40±0.02	0.41±0.07	0.18±0.05	90.01±1.89	1.16±0.05	0.70±0.05	0.63±0.06
NN-CFQP	90.01±0.51	0.39±0.01	0.02±.009	0.03±.009	89.95±1.62	1.54±0.12	0.12±0.04	0.12±0.03
	MEPS				GOV			
	Coverage	Length	KS (lo)	KS(hi)	Coverage	Length	KS (lo)	KS(hi)
Ln-CQR	89.92±0.66	0.66±0.01	0.09±0.03	0.33±0.05	90.00±0.19	0.79±.002	0.26±.014	0.44±0.02
Ln-CFQP	89.99±0.69	0.66±0.01	0.03±0.01	0.03±0.01	90.02±0.19	0.78±.002	0.05±0.01	0.04±0.01
RF-CQR	90.07±0.65	0.38±.009	0.19±0.02	0.30±0.03	90.03±0.17	0.61±.002	0.29±0.01	0.28±0.02
RF-CFQP	90.38±0.60	0.39±0.01	0.02±0.01	0.03±0.01	90.03±0.17	0.62±.002	0.05±0.01	0.04±0.01
NN-CQR	89.95±0.68	0.37±0.04	0.24±0.09	0.37± 0.06	90.01±0.19	0.58±0.01	0.28±0.03	0.32±0.04
NN-CFQP	89.97±0.61	0.37±0.04	0.03±0.01	0.04±0.01	90.01±0.18	0.59±0.01	0.05±0.01	0.05±0.01

Table 1: Results reported on test set of 200 repeated experiments with $\alpha = 0.1$. Ln, RF, and NN denote the linear, random forest, as well as neural network quantile regression models. Our methods are shown in bold.

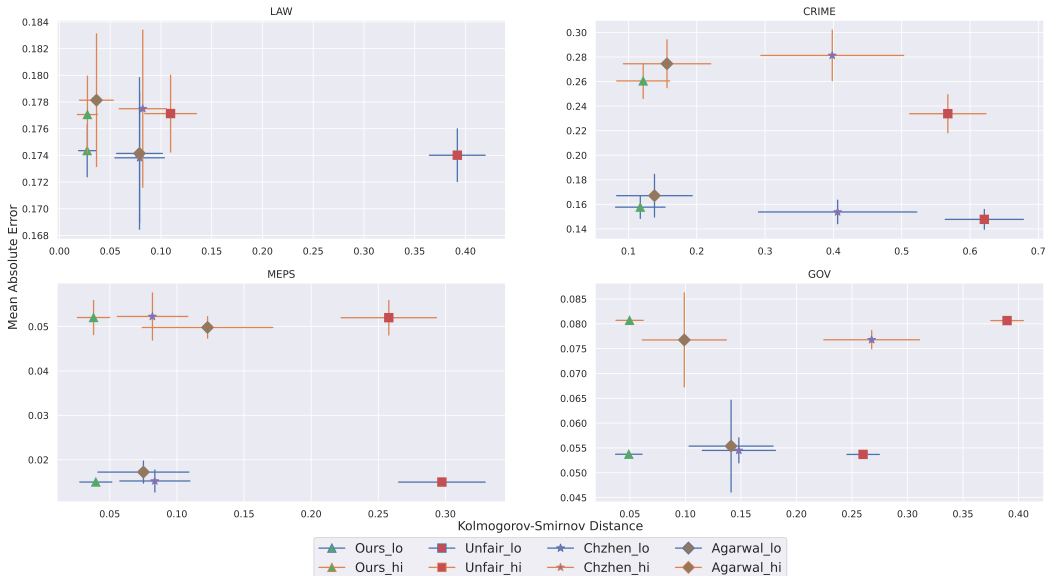

Figure 2: Results for estimating the lower (α_{lo}) and upper (α_{hi}) quantiles using some state-of-the-art DP-fairness requirement methods on all the datasets. We present the MAE and KS of lower quantile estimation, as well as upper quantile estimation. We applied linear quantile models in this comparison.

References

[1] A. Agarwal, M. Dudík, and Z. S. Wu. Fair regression: Quantitative definitions and reduction-based algorithms. In *International Conference on Machine Learning*, page 120–129. PMLR, 2019.

[2] M. Agueh and G. Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011. doi: 10.1137/100805741.

[3] R. Berk, H. Heidari, S. Jabbari, M. Joseph, M. Kearns, J. Morgenstern, S. Neel, and A. Roth. A convex framework for fair regression. *arXiv preprint arXiv:1706.02409*, 2017.

[4] M. J. Budig and M. J. Hodges. Statistical models and empirical evidence for differences in the motherhood penalty across the earnings distribution. *American Sociological Review*, 79(2):358–364, Mar 2014. doi: 10.1177/0003122414523616. URL `https://journals.sagepub.com/doi/full/10.1177/0003122414523616?casa_token=Q8sV04Yv5_IAAAAA:ve8iQgSXkPXnVQ8cGZjuLZWT4MIHEYCb-6Y2q9Zh8DghPUSlvpZpFNyXXfpte3qWka1MJl75gjPUPvQ`.

[5] F. Calmon, D. Wei, B. Vinzamuri, K. Natesan Ramamurthy, and K. R. Varshney. Optimized pre-processing for discrimination prevention. *Advances in neural information processing systems*, 30, 2017.

[6] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey. *Graphical methods for data analysis*. Chapman and Hall/CRC, 2018.

[7] C. Cheng and E. Parzen. Unified estimators of smooth quantile and quantile density functions. *Journal of statistical planning and inference*, 59(2):291–307, 1997.

[8] J. Cho, G. Hwang, and C. Suh. A fair classifier using kernel density estimation. *Advances in Neural Information Processing Systems*, 33:15088–15099, 2020.

[9] E. Chzhen and N. Schreuder. A minimax framework for quantifying risk-fairness trade-off in regression. *arXiv:2007.14265 [math, stat]*, Jan 2022. URL `http://arxiv.org/abs/2007.14265`. arXiv: 2007.14265.

[10] E. Chzhen, C. Denis, M. Hebiri, L. Oneto, and M. Pontil. Fair regression with wasserstein barycenters. *Advances in Neural Information Processing Systems*, 33:7321–7331, 2020. URL `https://proceedings.neurips.cc/paper/2020/hash/51cdbd2611e844ece5d80878eb770436-Abstract.html`.

[11] S. Corbett-Davies and S. Goel. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*, 2018.

[12] P. Czarnowska, Y. Vyas, and K. Shah. Quantifying social biases in nlp: A generalization and empirical comparison of extrinsic fairness metrics. *Transactions of the Association for Computational Linguistics*, 9: 1249–1267, 2021.

[13] L. Ding, D. Yu, J. Xie, W. Guo, S. Hu, M. Liu, L. Kong, H. Dai, Y. Bao, and B. Jiang. Word embeddings via causal inference: Gender bias reducing and semantic information preserving. *arXiv preprint arXiv:2112.05194*, 2021.

[14] M. Falk. Relative deficiency of kernel type estimators of quantiles. *The Annals of Statistics*, pages 261–268, 1984.

[15] A. for. Medical expenditure panel survey home, 2020. URL `https://meps.ahrq.gov/mepsweb/index.jsp`.

[16] S. Gallón, J.-M. Loubes, and E. Maza. Statistical properties of the quantile normalization method for density curve alignment. *Mathematical biosciences*, 242(2):129–142, 2013. doi: 10.1016/j.mbs.2012.12.007.

[17] J. García, P. J. Hernández, and A. López-Nicolás. How wide is the gap? an investigation of gender wage differences using quantile regression. *Empirical Economics*, 26(1):149–167, Mar 2001. doi: 10.1007/s001810000050. URL `https://link.springer.com/article/10.1007/s001810000050`.

[18] J. Gardeazabal and A. Ugidos. Gender wage discrimination at quantiles. *Journal of Population Economics*, 18(1):165–179, Mar 2005. doi: 10.1007/s00148-003-0172-z. URL `https://link.springer.com/article/10.1007/s00148-003-0172-z`.

[19] T. L. Gouic, J.-M. Loubes, and P. Rigollet. Projection to fairness in statistical learning. *arXiv:2005.11720 [cs, math, stat]*, Jun 2020. URL `http://arxiv.org/abs/2005.11720`. arXiv: 2005.11720.

[20] X. He, P. Ng, and S. Portnoy. Bivariate quantile smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):537–550, 1998.

[21] R. Jiang, A. Pacchiano, T. Stepleton, H. Jiang, and S. Chiappa. Wasserstein fair classification. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, page 862–872. PMLR, Aug 2020. URL https://proceedings.mlr.press/v115/jiang20a.html.

[22] J. L. Julia Angwin. Machine bias, May 2016. URL https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[23] A. Killewald and J. Bearak. Is the motherhood penalty larger for low-wage women? a comment on quantile regression. *American Sociological Review*, 79(2):350–7, 2014. URL http://asr.sagepub.com/content/79/2/350.short.

[24] R. Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005. doi: 10.1017/CBO9780511754098.

[25] R. Koenker and G. Bassett. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[26] R. Koenker, P. Ng, and S. Portnoy. Quantile smoothing splines. *Biometrika*, 81(4):673–680, 1994.

[27] M. J. Kusner, J. Loftus, C. Russell, and R. Silva. Counterfactual fairness. *Advances in neural information processing systems*, 30, 2017.

[28] E. L. Lehmann, J. P. Romano, and G. Casella. *Testing statistical hypotheses*, volume 3. Springer, 2005.

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

[30] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, Jul 2018. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2017.1307116.

[31] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, Jul 2018. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2017.1307116.

[32] Y. Li, Y. Liu, and J. Zhu. Quantile regression in reproducing kernel hilbert spaces. *Journal of the American Statistical Association*, 102(477):255–268, Mar 2007. ISSN 0162-1459. doi: 10/b8mq5k.

[33] N. Meinshausen and G. Ridgeway. Quantile regression forests. *Journal of Machine Learning Research*, 7 (6), 2006.

[34] S. Milli, J. Miller, A. D. Dragan, and M. Hardt. The social cost of strategic classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 230–239, 2019.

[35] L. Oneto, M. Donini, and M. Pontil. General fair empirical risk minimization. *arXiv:1901.10080 [cs, stat]*, Dec 2019. URL http://arxiv.org/abs/1901.10080. arXiv: 1901.10080.

[36] M. Pietrosanu, J. Gao, L. Kong, B. Jiang, and D. Niu. Advanced algorithms for penalized quantile and composite quantile regression. *Computational Statistics*, 36(1):333–346, Mar 2021. ISSN 0943-4062, 1613-9658. doi: 10.1007/s00180-020-01010-1.

[37] D. Plecko and N. Meinshausen. Fair data adaptation with quantile preservation. *Journal of Machine Learning Research*, 21(242):1–44, 2020.

[38] D. Plečko, N. Bennett, and N. Meinshausen. fairadapt: Causal reasoning for fair data pre-processing. *arXiv preprint arXiv:2110.10200*, 2021.

[39] S. Portnoy and R. Koenker. The gaussian hare and the laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4):279–300, 1997.

[40] M. Redmond and A. Baveja. A data-driven software tool for enabling cooperative information sharing among police departments. *European Journal of Operational Research*, 141(3):660–678, 2002.

[41] Y. Romano, R. Barber, C. Sabatti, and E. Candès. With malice towards none: Assessing uncertainty via equalized coverage. *ArXiv*, 2019. doi: 10.1162/99608f92.03f00592.

[42] Y. Romano, E. Patterson, and E. Candes. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://papers.nips.cc/paper/2019/hash/5103c3584b063c431bd1268e9b5e76fb-Abstract.html.

[43] I. Takeuchi, Q. V. Le, T. D. Sears, and A. J. Smola. Nonparametric quantile estimation. *Journal of Machine Learning Research*, 7(45):1231–1264, 2006. URL `http://jmlr.org/papers/v7/takeuchi06a.html`.

[44] J. W. Taylor. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*, 19(4):299–311, 2000.

[45] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world*. Springer, New York, 2005. ISBN 978-0-387-00152-4.

[46] V. Vovk, I. Nouretdinov, and A. Gammerman. On-line predictive linear regression. *The Annals of Statistics*, pages 1566–1590, 2009.

[47] L. F. Wightman. Lsac national longitudinal bar passage study. lsac research report series. 1998.

[48] D. Yang, J. Lafferty, and D. Pollard. Fair quantile regression. *arXiv preprint arXiv:1907.08646*, 2019.

[49] S. Yuan. Random gradient boosting for predicting conditional quantiles. *Journal of Statistical Computation and Simulation*, 85(18):3716–3726, 2015.

[50] M. B. Zafar, I. Valera, M. Gomez Rodriguez, and K. P. Gummadi. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th international conference on world wide web*, pages 1171–1180, 2017.

[51] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork. Learning fair representations. In *International conference on machine learning*, pages 325–333. PMLR, 2013.

[52] Z. Zhang and H.-G. Müller. Functional density synchronization. *Computational Statistics & Data Analysis*, 55(7):2234–2249, 2011. doi: 10.1016/j.csda.2011.01.007.

[53] K. Q. Zhou and S. L. Portnoy. Direct use of regression quantiles to construct confidence sets in linear models. *The Annals of Statistics*, 24(1):287–306, 1996.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See Section 8
   (c) Did you discuss any potential negative societal impacts of your work? [No] We currently don't foresee the potential negative societal impacts of our work, however, we anticipate this method to be used in the future by practitioners to real-life scenarios involving job markets, medical science, etc., and to potentially help making decisions that would protect the vulnerable and propel social harmony.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes]

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]