Supplementary Material for " Conformalized Fairness via Quantile Regression"

A Conditions for Proposition 2.

Condition (A1). The quantile density functions $Q'_{s'}, s' = 1, \dots, K$, 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 $0 < c_0 \leq \inf_{1 \leq s' \leq K} \tau_{s'} \leq \sup_{1 \leq s' \leq K} \tau_{s'} < C_0 < \infty, s' = 1, \dots, K$.

Remark 1. Conditions (A1) - (A3) guarantee the existence of a strong approximation of the empirical quantile process by a sequence of weighted Brownian bridges as established in [6]. Condition (A3) posited on kernel functions assures that the integral transform $\hat{F}^{-1} \mapsto \hat{Q}$ possesses good approximation properties for smooth functions, and it is shown that (A3) holds for any difference kernel $\mathrm{d}_t K_h(u,t) = h^{-1} k\left((u-t)/h\right) \mathrm{d}t$ with a vanishing bandwidth h. for example, the gaussian density $K(u) = \exp(-u^2/2h^2)$ [3, 16, 17] or the triangular density function $K(u) = (1 - |u|/h) I(|u|/h \leq 1)$ with a vanishing bandwidth h_n.

B Proofs related to CFQP

In this section, we prove the validity of the CFQP prediction intervals described in Section 4 of the main paper. First, we recap some results on distribution-free order statistics from Romano et al. [11].

The quantile function Q of a random variable Z, with cumulative distribution function $F(z) := P\{Z \leq z\}$, is defined by the equivalence

$$Q(\alpha) \leq z \quad \text{if and only if} \quad \alpha \leq F(z)$$

for all $\alpha \in (0, 1)$ and $z \in \mathbb{R}$. And less standardly, the right quantile function R of the random variable Z is defined by the equivalence $F^-(z) \leq \alpha$ if and only if $z \leq R(\alpha)$, where $F^-(z) := F(z-) = P\{Z < z\}$. The quantile functions have the explicit formulas

$$Q(\alpha) = \inf\{z \in \mathbb{R} : \alpha \leq F(z)\}, \quad R(\alpha) = \sup\left\{z \in \mathbb{R} : F^-(z) \leq \alpha\right\}.$$

As a special case, the empirical quantile function $\hat{Q}_n$ of random variables $Z_1, \dots, Z_n$ is the quantile function with respect to the empirical CDF $\hat{F}_n(z) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{Z_i \leq z\}$. Likewise, the right empirical quantile function $\hat{R}_n$ of $Z_1, \dots, Z_n$ is the right quantile function with respect to $\hat{F}_n^-(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{Z_i < z\}$. They have the explicit formulas

$$\hat{Q}_n(\alpha) = Z_{(\lceil \alpha n \rceil)}, \quad \hat{R}_n(\alpha) = Z_{(\lfloor \alpha n \rfloor + 1)},$$

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

where $Z_{(k)}$ denotes the k th smallest value in $Z_1, \ldots, Z_n$. Several variants of the following lemmas are showed in Vovk et al. [13],Vovk et al. [14] and Romano et al. [12]. We restate them here for clarification.

Lemma 1 (Quantiles and exchangeability)**.** Suppose $Z_1, \ldots, Z_n$ are exchangeable random variables. For any $\alpha \in (0,1)$,

$$P\left\{Z_n \leq \hat{Q}_n(\alpha)\right\} \geq \alpha.$$

Moreover, if the random variables $Z_1, \ldots, Z_n$ are almost surely distinct, then also

$$P\left\{Z_n \leq \hat{Q}_n(\alpha)\right\} \leq \alpha + \frac{1}{n}.$$

Here the probabilities are taken over all the variables $Z_1, \ldots, Z_n$.

Lemma 2 (Inflation of quantiles)**.** Suppose $Z_1, \ldots, Z_{n+1}$ are exchangeable random variables. For any $\alpha \in (0,1)$,

$$P\left\{Z_{n+1} \leq \hat{Q}_n\left(\left(1+\frac{1}{n}\right)\alpha\right)\right\} \geq \alpha.$$

Moreover, if the random variables $Z_1, \ldots, Z_{n+1}$ are almost surely distinct, then also

$$P\left\{Z_{n+1} \leq \hat{Q}_n\left(\left(1+\frac{1}{n}\right)\alpha\right)\right\} \leq \alpha + \frac{1}{n+1}.$$

Applying the previous auxiliary lemmas, we can prove the validity of the CFQP prediction intervals described in Section 4.

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

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

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

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

Proof of Theorem 1. Conditionally on the proper training set. Denote by E_{n+1} the conformity score

$$E_i := \max\left\{\hat{g}_{\alpha_{\mathrm{lo}}}\left(\tilde{X}_i\right) - Y_i, Y_i - \hat{g}_{\alpha_{\mathrm{hi}}}\left(\tilde{X}_i\right)\right\}$$

at the test point $\left(\tilde{X}_{n+1}, Y_{n+1}\right)$. By the construction of the prediction interval, we have

$$Y_{n+1} \in C\left(\tilde{X}_{n+1}\right) \quad \text{if and only if} \quad E_{n+1} \leq Q_{1-\alpha}\left(E, \mathcal{I}_2\right),$$

and thus,

$$P\left\{Y_{n+1} \in C\left(\tilde{X}_{n+1}\right) \mid \left(\tilde{X}_i, Y_i\right) : i \in \mathcal{I}_1\right\} = P\left\{E_{n+1} \leq Q_{1-\alpha}\left(E, \mathcal{I}_2\right) \mid \left(\tilde{X}_i, Y_i\right) : i \in \mathcal{I}_1\right\} \tag{1}$$

Since the original pairs $\left(\tilde{X}_i, Y_i\right)$ are exchangeable, so are the calibration variables E_i for $i \in \mathcal{I}_2$ and $i = n+1$. Thus, by Lemma 2 on inflated empirical quantiles,

$$P\left\{E_{n+1} \leq Q_{1-\alpha}\left(E, \mathcal{I}_2\right) \mid \left(\tilde{X}_i, Y_i\right) : i \in \mathcal{I}_1\right\} \geq 1-\alpha, \tag{2}$$

and, under the additional assumption that the E_i 's are almost surely distinct,

$$P\left\{E_{n+1} \leq Q_{1-\alpha}\left(E, \mathcal{I}_2\right) \mid \left(\tilde{X}_i, Y_i\right) : i \in \mathcal{I}_1\right\} \leq 1-\alpha + \frac{1}{|\mathcal{I}_2|+1} \tag{3}$$

The exact coverage result is derived by taking expectations over the proper training set in 1, 2, and 3. $\square$

Next, we likewise prove the validity of the extended CFQP prediction intervals that control the left and right tails independently.

Corollary 1. Define the prediction interval

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

where $Q_{1-\alpha_{\rm lo}}\left(E_{\rm lo}, \mathcal{I}_2\right)$ is the $(1-\alpha_{\rm lo})$-th empirical quantile of $\{\hat{g}_{\alpha_{\rm lo},i} - Y_i : i \in \mathcal{I}_2\}$ and $Q_{1-\alpha_{\rm hi}}\left(E_{\rm hi}, \mathcal{I}_2\right)$ is the $(1-\alpha_{\rm hi})$-th empirical quantile of $\{Y_i - \hat{g}_{\alpha_{\rm hi},i} : i \in \mathcal{I}_2\}$. If the samples $\left(\tilde{X}_{n+1}, Y_i\right), i = 1, \ldots, n+1$ are exchangeable, then

$$P\left\{Y_{n+1} \geq \hat{g}_{\alpha_{\rm lo}}\left(\tilde{X}_{n+1}\right) - Q_{1-\alpha_{\rm lo}}\left(E_{\rm lo}, \mathcal{I}_2\right)\right\} \geq 1 - \alpha_{\rm lo}, \tag{4}$$

and
$$P\left\{Y_{n+1} \leq \hat{g}_{\alpha_{\rm hi}}\left(\tilde{X}_{n+1}\right) + Q_{1-\alpha_{\rm hi}}\left(E_{\rm hi}, \mathcal{I}_2\right)\right\} \geq 1 - \alpha_{\rm hi}. \tag{5}$$

Consequently, we also have $P\left\{Y_{n+1} \in C\left(\tilde{X}_{n+1}\right)\right\} \geq 1 - \alpha$ assuming $\alpha = \alpha_{\rm lo} + \alpha_{\rm hi}$.

Proof. The two events inside the probabilities (4) as well as (5) are equivalent to $\hat{g}_{\alpha_{\rm lo}}\left(X_{n+1}\right) - Y_{n+1} \leq Q_{1-\alpha_{\rm lo}}\left(E_{\rm lo}, \mathcal{I}_2\right)$ and $Y_{n+1} - \hat{g}_{\alpha_{\rm hi}}\left(X_{n+1}\right) \leq Q_{1-\alpha_{\rm hi}}\left(E_{\rm hi}, \mathcal{I}_2\right)$, respectively. The results are derives by applying Lemma 2 twice, in the same manner as in the proof of Theorem 1. □

C Proofs related to DP

Proof of Proposition 2. By Theorem 2.1(2) in Cheng and Parzen [3], we have

$$\sup_{t\in[0,1]}\left|\hat{Q}_{2,q_\alpha|s'}(t) - Q_{q_\alpha|s'}(t)\right| = O_p\left(N^{-1/2} + N^{-\rho}\right) = O_p\left(N^{-1/2}\right), \tag{6}$$

for each k, as $\rho > 1/2$. according to Assumption (A1) , conditions (Q1)-(Q3) in Cheng and Parzen [3] are satisfied, and since we choose the kernel function K_h with the properties in (A3), conditions (K1)-(K3) in Cheng and Parzen [3] are ensured. Moreover, the extension from $[a, b] \subset (0, 1)$ to $[0, 1]$ is made possible by Assumption (A1). At last, assumption (A2) and the fact that the bound in equation (2.7) in Cheng and Parzen [3] is a universal bound which does not depend on s' allow the extension from Eq. 6 to

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

□

Before showing the exact DP guarantee, we utilize Lemma E.1. (stated in Lemma 3) from Chzhen and Schreuder [4] and references therein, where rigorous proofs are given.

Lemma 3. Let $V_1, \ldots, V_n, V_{n+1}, n \geq 1$ be exchangeable real-valued random variables and U distributed uniformly on $[0, 1]$ be independent from $V_1, \ldots, V_n, V_{n+1}$, then the constructed location statistic

$$T\left(V_1, \ldots, V_n, V_{n+1}, U\right) = \frac{1}{n+1}\left(\sum_{i=1}^{n}\mathbb{1}\left\{V_i < V_{n+1}\right\} + U \cdot \left(1 + \sum_{i=1}^{n}\mathbb{1}\left\{V_i = V_{n+1}\right\}\right)\right)$$

is distributed uniformly on $[0, 1]$.

The proof of Theorem 2 for quantile DP is a direct adaptation of demographic parity guarantee from Chzhen and Schreuder [4] for the mean regression.

Proof of Theorem 2. To prove the claimed DP guarantee for fixed quantile level $\alpha \in \{\alpha_{\rm lo}, \alpha_{\rm hi}\}$, we will show that the Kolmogorov-Smirnov distance between $\nu_{\hat{g}_\alpha|s}$ and $\nu_{\hat{g}_\alpha|s'}$ equals to zero for any $s \neq s' \in [K]$.

Note that, according to the formulation of $\hat{g}_\alpha$ in Eq. (12), we have for any $(x, s) \in \mathbb{R}^p \times [K]$,

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

Denote by $\hat{Q}(t) = \sum_{s'=1}^{K} \hat{p}_{s'} \hat{Q}_{2,q_\alpha|s'}$. Note that we use the training set to estimate the location statistic for the new test point, $\hat{Q}(t)$ is independent from $\hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x, s)$ for each $s \in [K]$.

Since the test point belongs to group $S = s$ for some fixed $s \in [K]$ and, for all $i = 1, \ldots, |\mathcal{I}_1^s|$, set $V_i = \tilde{q}^s_{1,i}$ with $V_{N_s+1} \overset{d}{=} (\tilde{q}_\alpha(x, s))$ independent from $(V_i)_{i=1,\ldots,|\mathcal{I}_1^s|}$. Since the random variables $V_1, \ldots, V_{N_s}, V_{N_s+1}$ are exchangeable (more ideally, independent), Lemma 3 implies that for all $s \in [K]$, the location statistic $\hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x, s)$ is distributed uniformly on $[0, 1]$. Thus for all $s, s' \in [K]$, we have

$$\begin{aligned}
&\mathrm{KS}\left(\nu_{\hat{g}_\alpha|s}, \nu_{\hat{g}_\alpha|s'}\right) \\
=&\sup_{t\in\mathbb{R}} \left|P(\hat{g}_\alpha \le t \mid S = s) - P\left(\hat{g}_\alpha \le t \mid S = s'\right)\right| \\
=&\sup_{t\in\mathbb{R}} \left|P\left(\hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x, s) \le \hat{Q}^{-1}(t) \mid S = s\right) - P\left(\hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x, s) \le \hat{Q}^{-1}(t) \mid S = s'\right)\right| \\
=&\sup_{t\in\mathbb{R}} \left|E\left[\hat{Q}^{-1}(t) \mid S = s\right] - E\left[\hat{Q}^{-1}(t) \mid S = s'\right]\right| = 0.
\end{aligned}$$

The first equality uses the definition of Eq. (9); the second uses the fact that $\hat{Q}$ is monotone by construction [4]; finally since the independence of $\hat{Q}$ is independent from $\hat{F}_{1,q_\alpha|s} \circ \tilde{q}_\alpha(x, s)$ conditionally on $S = s$ for any $s \in [K]$, also $\hat{Q}$ remains independent from S. The exact DP is concluded. □

D Experiments

D.1 Data Description & Pre-processing

1. *Law School (LAW)*[15]: 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).
2. *Community&Crime (CRIME)*[10]: This dataset contains socio-economic, law enforcement, and crime data about communities in the US with 1,994 examples. The task is to predict the number of violent crimes per 100,000 population with race as the sensitive attribute.
3. *MEPS 2016 (MEPS)*[1, 12]: 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 including age, marital status, race, poverty status, health status, health insurance type, and more. There are 15,656 examples on $p = 41$ features with race as the sensitive attribute (nonwhite vs. white).
4. *Government Salary (GOV)*[8, 9]: The government salary dataset (available in R package "fairadapt") is collected from the 2018 American Community Survey by the US Census Bureau. The yearly salary for over 200,000 examples is the response variable, and employee race (7 categories) is identified as the sensitive attribute.

In order to smoothly run the regression model, several preprocessing steps are utilized before running the model. For example, in the CRIME dataset, we impute the missing value by mean, and the race feature is created by the maximum value of the four races. The clean data are uploaded to the GitHub repo for future research.

D.2 Adaptation of other algorithms in [2, 5]

First, our approach is built upon the one proposed by Chzhen et al. [5], we incorporated the kernel smoothing procedure in quantile estimation and applied the local linear smoothing method provided in *NumPy* [7] which shows its functionality when there are subgroups of small sample sizes, especially for the *CRIME* dataset. It is also feasible to compute Eq. (8) via some global kernels, such as the gaussian density $K(u) = \exp(-u^2/2h^2)$ [3, 16, 17] or the triangular density function $K(u) = (1 - |u|/h)I(|u|/h \leq 1)$ with a vanishing bandwidth h_n. For the bandwidth selection, there is a publicly available R package *lokern* with the global bandwidth choice.

Second, we adjusted the reduction-based approach in Agarwal et al. [2] in several respects: we rescale and discretize the responses, and modify their algorithm by replacing the loss function l by ρ_α of Eq. (4) in our paper. It is worth mentioning that the reduction-based approach is sensitive to the hyperparameters: discretization parameter N and slack $\hat{\varepsilon}_a$ in training. We used logistic regression and SVM classifiers in tuning.

D.3 Additional experiment results

We show the additional results using quantile random forest and quantile neural network for the comparison of our post-processing fairness adjustment procedure in figures 1 and 2.

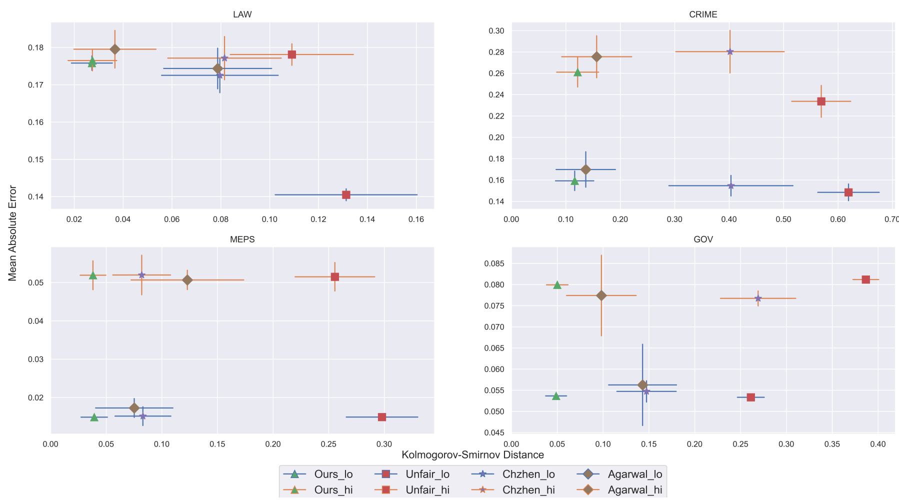

Figure 1: Results using Quantile random forest for estimating the lower (α_{lo}) and upper (α_{hi}) quantiles using some state-of-the-art DP-fairness requirement methods on all the datasets. ‘Unfair’, ‘Chzhen’, and ‘Agarwal’ stand for the quantile model without fairness adjustment, barycenter method [5] and reduction-based algorithm [2] respectively. We present the MAE and KS of lower quantile estimation, as well as upper quantile estimation. We set ‘n estimator’ to 50.

	LAW			
	Coverage	Length	KS(lo)	KS(hi)
Ln-CFQP (with both smoothing)	90.02±0.51	0.46±.004	0.02±0.01	0.02±0.01
Ln-CFQP (with jittering)	89.99±0.49	0.39±.002	0.03±.008	0.03±0.01
Ln-CFQP (with kernel smoothing)	89.99±0.50	0.38±.002	0.03±0.009	0.03±0.01
Ln-CFQP (without smoothing)	89.96±0.49	0.40±.002	0.04±0.01	0.03±0.01
	CRIME			
	Coverage	Length	KS(lo)	KS(hi)
Ln-CFQP (with both smoothing)	90.44±1.84	1.64±0.05	0.11±0.03	0.12±0.04
Ln-CFQP (with jittering)	90.55±1.35	1.69±0.05	0.31±0.11	0.33±0.12
Ln-CFQP (with kernel smoothing)	90.58±1.36	1.69±0.05	0.25±0.13	0.29±0.12
Ln-CFQP (without smoothing)	90.53±1.27	1.69±0.05	0.36±0.12	0.40±0.12

Table 1: Ablation test results when removing either or both of the smoothing strategies. Our methods are shown in bold.

The three quantile models present consistent results, our post-processing based upon kernel smoothing outperforms the other two approaches, which is reflected in more KS value reduction. We also conducted a brief ablation study by testing the experimental results of removing either or both of the smoothing strategies in our CFQP approach. Results are presented in table 1. We found incorporating the jittering and kernel smoothing methods works better when the subgroups are unbalanced and there exist subgroups of small sample sizes, especially for the *CRIME* dataset. Finally, the computational time analysis is also incorporated in figure 3.

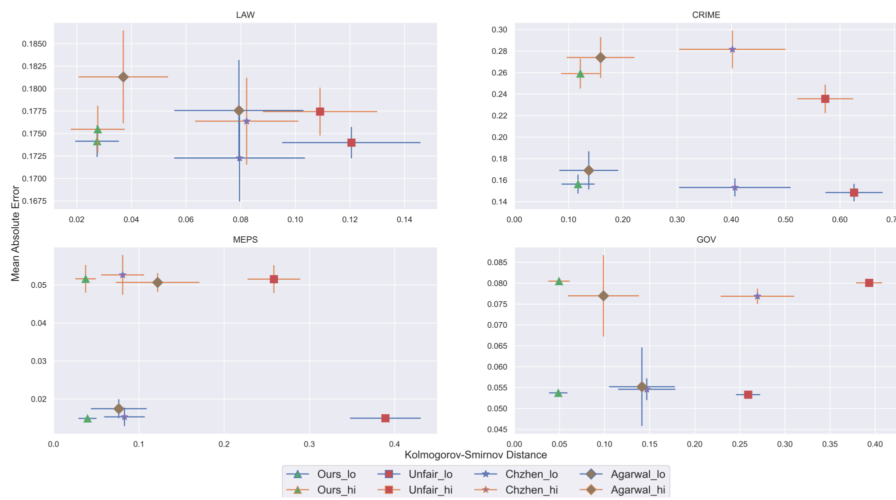

Figure 2: Results using Quantile neural network model for estimating the lower (α_{lo}) and upper (α_{hi}) quantiles using some state-of-the-art DP-fairness methods on all the datasets. 'Unfair', 'Chzhen', and 'Agarwal' stand for the quantile model without fairness adjustment, barycenter method [5] and reduction-based algorithm [2] respectively. We present the MAE and KS of lower quantile estimation, as well as upper quantile estimation.

In the current experiment, the kernel we used is the local linear one in defining the quantile functions of subgroups instead of the global kernels. When calculating the τ-th quantile x_τ of q_α using local linear smoothing, we choose a constant distance size h (kernel radius) and compute a weighted average for all data points that are closer to x_τ (the closer to x_τ points get higher weights). The time complexity for computing the local kernel smoother (Eq.(8)) is $\mathcal{O}(1)$, while if we applied the global kernels in Eq.(8), $\mathcal{O}(n)$ time would cost.

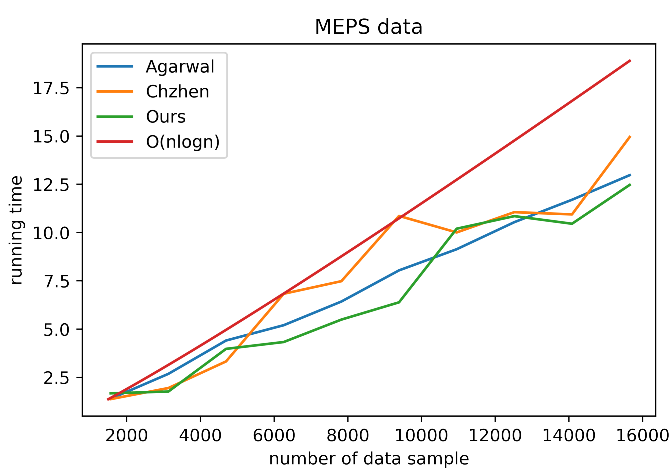

Figure 3: Empirical running times of methods used for experimental comparisons. We utilized the linear quantile model on the MEPS dataset. For methods of "Chzhen" and "Agarwal", normalizing factors are applied to present the graph.

For the CFQP method, The steps determining the time complexity of Algorithms 1 and 2 reside in the following two parts:

1. The for-loop where we perform a post-processing which takes $\sum_{s' \in [K]} \mathcal{O}\left(N_{s'} \log N_{s'}\right)$ time, as we need to sort the grouped samples;
2. The evaluation of $\hat{g}_\alpha$ on a new point (x, s) is performed in $\max_{s' \in [K]} \mathcal{O}(\log N_s)$ time as it involves locating $\hat{g}_\alpha$ in a sorted array.

References

[1] Medical expenditure panel survey. URL `https://meps.ahrq.gov/mepsweb/`.

[2] 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.

[3] 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.

[4] E. Chzhen and N. Schreuder. A minimax framework for quantifying risk-fairness trade-off in regression. *The Annals of Statistics*, 50(4):2416–2442, 2022.

[5] 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.

[6] M. Csorgo and P. Revesz. Strong approximations of the quantile process. *The Annals of Statistics*, 6(4):882–894, 1978. ISSN 00905364.

[7] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2.

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

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

[10] 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.

[11] 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.

[12] Y. Romano, E. Patterson, and E. Candes. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

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

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

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

[16] S.-S. Yang. A smooth nonparametric estimator of a quantile function. *Journal of the American Statistical Association*, 80(392):1004–1011, 1985. ISSN 0162-1459. doi: 10.2307/2288567.

[17] 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.