Robust Quantile Regression Under Unreliable Data

Yoshifumi Shoji, Masahiro Yukawa
Department of Electronics and Electrical Engineering, Keio University, Yokohama, Japan
E-mail: shoji.yoshifumi@ykw.elec.keio.ac.jp, yukawa@elec.keio.ac.jp

***Abstract*—This paper addresses the quantile regression task when some non-negligible portion of data are corrupted by accidental factors such as temporary sensor malfunctions. Here, the task is to find the empirical quantile of the "reliable" data with the "unreliable" ones excluded. For this task, we propose the MC-pinball loss which is the composition of the minimax concave (MC) penalty and the pinball loss. The simulation results show that the proposed approach yields reasonable estimates of the true quantile. A potential benefit of the proposed approach is also shown with respect to the parameter tuning.**

I. Introduction

Quantile regression [1] has been studied for more than half a century, and it has been applied to various problems, including wind power generation [2], [3], power forecast [4], wage estimation [5], [6], gold price estimation [7], to name just a few. It still remains an active research topic in many areas including signal processing [8] and statistics [9].

We consider the situation where the input vectors $(\boldsymbol{x}_i)_{i=1}^m$ and the outputs $(y_i)_{i=1}^m$ are given. The key assumption here is that a non-negligible portion of the outputs are corrupted by accidental factors such as temporary sensor malfunctions, measurement/transmission errors, and human errors. Those outputs will be referred to as *unreliable* data. The rest of the outputs are *reliable* data assumed to be modeled well by means of linear estimation. As such, we define the residual vectors

$$\boldsymbol{\epsilon} := [\epsilon_1, \epsilon_2, \cdots, \epsilon_m]^\top := \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}_\star \in \mathbb{R}^m, \quad (1)$$

where $\boldsymbol{y} := [y_1, y_2, \cdots, y_m]^\top \in \mathbb{R}^m$, $\boldsymbol{X} := [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \cdots \ \boldsymbol{x}_m]^\top \in \mathbb{R}^{m\times n}$, and $\boldsymbol{\theta}_\star \in \mathbb{R}^n$ is the 'true' regression vector.

The deterministic $\boldsymbol{x}_i$s, y_is, and ϵ_is are samples of random vector X and rondom variables Y and ε, respectively. Here, ε obeys a mixture of distributions of which the probability density function is given by

$$h(\epsilon) := (1-\beta)f(\epsilon) + \beta g(\epsilon), \ \epsilon \in \mathbb{R}, \quad (2)$$

where f is the density of ε for reliable data, and g is that for unreliable data with its proportion $\beta \in [0,1]$. The set of ϵ_is for reliable data is denoted by $\mathcal{D}_\mathrm{r} = (\epsilon_i)_{i\in\mathcal{I}_\mathrm{r}} (\subset \mathcal{D} := (\epsilon_i)_{i\in\mathcal{I}})$, where $\mathcal{I}_\mathrm{r} \subset \mathcal{I} := \{1, 2, \cdots, m\}$. The conditional density of Y given $X = \boldsymbol{x}$ is given by $h_Y(y \mid X = \boldsymbol{x}) = h(y - \boldsymbol{x}^\top\boldsymbol{\theta}_\star)$, where $y - \boldsymbol{x}^\top\boldsymbol{\theta}_\star = \epsilon$. We also define $f_Y(y \mid X = \boldsymbol{x}) = f(y - \boldsymbol{x}^\top\boldsymbol{\theta}_\star)$ and $g_Y(y \mid X = \boldsymbol{x}) = g(y - \boldsymbol{x}^\top\boldsymbol{\theta}_\star)$.

To clarify the concept, the situation under consideration is illustrated in Fig. 1. The reliable data are concentrated, while the unreliable data reside in those regions which are apart from the region of reliable data. Suppose, for instance, that one is interested in the traffic of wireless communication, or more specifically, one would like to know such an interval in which the "actual" traffic falls with a prespecified probability. In this context, each "reliable" datum y_i is the actual traffic given $\boldsymbol{x}_i$, and ϵ_i is the mismatch from its 'best' linear approximation in some sense. In this case, y_i is regarded as a sample of Y generated from $f_Y(Y \mid X = \boldsymbol{x}_i)$. The "unreliable" data, on the other hand, deviate from the actual traffic, and those data may cause fatal errors in quantile regression.

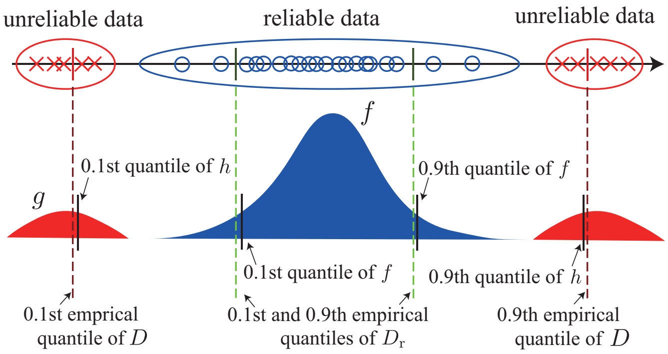


Fig. 1: An example of density of ε. Circle/cross indicates $\epsilon_i (= y_i - \boldsymbol{x}_i^\top\boldsymbol{\theta}_\star)$ for reliable/unreliable data.

Keeping the above arguments in mind, the desirable interval would be estimated by finding the empirical quantiles using the dataset that solely contains the "reliable" data. Since separating those reliable and unreliable data is a nontrivial task, the goal of the current study is to build a loss function which is insensitive to the "unreliable" data so that its minimizer gives a reasonable estimate of the empirical quantiles of the "reliable" data. The so-called pinball loss (an asymmetric least absolute deviation) is typically used for quantile regression, and the question address in this work is how to make it insensitive to the "unreliable" data.

In this paper, we first consider the use of Moreau enhancement of the pinball loss function, motivated by the fact that the Moreau enhancement of the absolute-value function has been used successfully for robust signal recovery [10]. It then turns out to be undesirable for the present task. We therefore propose an alternative approach based on the MC-pinball loss which is a composition of the pinball loss with the minimax concave (MC) penalty [11], [12]. The proposed loss has a couple of nice properties leading to insensitivity to the unreliable data. Numerical studies are presented with f being the Gaussian

This work was partially supported by JST SICORP Grant Number JPMJSC20C6, Japan.

distribution and with g being three different distributions: the uniform distribution, the gamma distribution, and a point mass (Dirac's delta function). We demonstrate that the proposed approach yields a good estimate of the true quantile in terms of system mismatch (the quantile estimation error) and coverage rate even for a large β up to 0.5, while the performance of the conventional approach based on the pinball loss degrades sharply as β increases slightly from zero.

II. Robust Quantile Regression Problem under Unreliable Data

Throughout the paper, let $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{R}_{++}$ denote the sets of nonnegative integers, real numbers, and strictly positive real numbers, respectively. Let X be a real-valued random vector of dimension $n \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}$, and Y be a real-valued random variable which is related to X. The αth conditional quantile function of Y given $X = \boldsymbol{x}$ is defined by $q_\alpha(\boldsymbol{x}) := \inf\{y \in \mathbb{R} : F(y \mid X = \boldsymbol{x}) \geq \alpha\}$, where X may be characterized by its (right-continuous) distribution function $F(y \mid X = \boldsymbol{x}) := \mathbb{P}\{Y \leq y \mid X = \boldsymbol{x}\}$ [1], [9].

In practical applications, the data could be changed into completely different values due to a variety of reasons during the data acquisition/transmission process. Such "unreliable" data do *not* obey the true distribution, and hence the empirical αth quantile using the whole dataset containing the unreliable data could be significantly different from of the data distribution. The task is therefore to estimate the αth empirical quantile of the set of "reliable" data, which is expected to be a reasonable estimate of the true quantile of $f_Y(y \mid X = \boldsymbol{x})$. Here, we assume implicitly that those unreliable data reside mostly in those regions which are apart from the bulk of reliable data, i.e., such unreliable data that are mixed up with the reliable one are supposed to give negligible impacts on the estimation. See Fig. 1.

The αth empirical quantile $\boldsymbol{x}^\top \hat{\boldsymbol{\theta}}_\alpha$ at $\boldsymbol{x}$ from the training dataset is obtained via minimization of the pinball loss function [1] $\sum_{i=1}^m \rho_\alpha(y_i - \boldsymbol{x}_i^\top \boldsymbol{\theta})$, where (see Fig. 2)

$$\rho_\alpha : \mathbb{R} \to [0, +\infty) : z \mapsto \begin{cases} \alpha z, & \text{if } z \geq 0, \\ -(1-\alpha)z, & \text{otherwise.} \end{cases} \tag{3}$$

In the present scenario, the empirical quantile obtained with the pinball loss would be an estimate of the quantile of $h_Y(y \mid X = \boldsymbol{x})$, while what is desired to estimate is the quantile of the density $f_Y(y \mid X = \boldsymbol{x})$ of the reliable data. In the following section, we present a modified loss function to seek for the empirical quantile of the reliable data.

III. Proposed Approach to Robust Quantile Regression

Under the implicit assumption stated in the previous section, the idea is to modify the pinball loss in such a way that it saturates when $|z|$ exceeds some thresholds. A similar idea has been studied in the context of robust signal recovery [10]

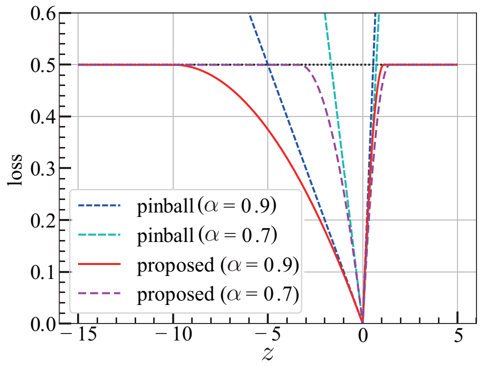


Fig. 2: Pinball and proposed losses ($\alpha = 0.7, 0.9$)

using the MC penalty [11], [12]

$$\phi_\gamma^{\mathrm{MC}} : \mathbb{R} \to [0, +\infty) : z \mapsto \begin{cases} |z| - \dfrac{1}{2\gamma} z^2, & \text{if } |z| < \gamma, \\ \dfrac{1}{2}\gamma, & \text{otherwise,} \end{cases} \tag{4}$$

where $\gamma \in \mathbb{R}_{++}$ is the saturation factor.

In fact, the MC penalty can be expressed as $\phi_\gamma^{\mathrm{MC}} = |\cdot| - {}^\gamma|\cdot|$. Here, for any convex function $f : \mathbb{R} \to \mathbb{R}$, its Moreau envelope of index $\gamma > 0$ is defined by

$${}^\gamma f(z) := \min_{x \in \mathbb{R}} \left(f(z) + \frac{1}{2\gamma}(z - x)^2 \right), \ z \in \mathbb{R}. \tag{5}$$

The function $\phi_\gamma^{\mathrm{MC}}$ has an enhanced sparsity-promoting property owing to the Moreau envelope ${}^\gamma|\cdot|$ of $|\cdot|$, and thus $\phi_\gamma^{\mathrm{MC}}$ is referred to as the Moreau enhancement [13] of $|\cdot|$.

We first consider a straightforward approach of applying the Moreau enhancement to the pinball loss. This approach, however, will turn out to be undesirable in the present context. Therefore, we propose an alternative approach by composing the two functions ρ_α and $\phi_\gamma^{\mathrm{MC}}$, and we finally compare it with the straightforward approach.

A. Moreau-enhanced Pinball Loss and Its Issues

The Moreau-enhanced pinball function is given by

$$\begin{aligned} \phi_{\alpha,\gamma} : \mathbb{R} \to [0, +\infty) : z \mapsto \rho_\alpha(z) - {}^\gamma\rho_\alpha(z) \\ = \begin{cases} \dfrac{\alpha^2}{2}\gamma, & \text{if } z \geq \alpha\gamma, \\ \alpha z - \dfrac{1}{2\gamma} z^2, & \text{if } 0 \leq z < \alpha\gamma, \\ -(1-\alpha)z - \dfrac{1}{2\gamma} z^2, & \text{if } -(1-\alpha)\gamma < z < 0, \\ \dfrac{(1-\alpha)^2}{2}\gamma, & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

The process of generating $\phi_{\alpha,\gamma}$ from $|\cdot|$ is indicated in Fig. 3 by the blue arrow. Different weights are given to each side of $|\cdot|$ to obtain the pinball loss ρ_α (bottom left). The Huberized pinball (bottom middle, the Moreau envelope of ρ_α) is subtracted from ρ_α to obtain $\phi_{\alpha,\gamma}$. It can be seen that

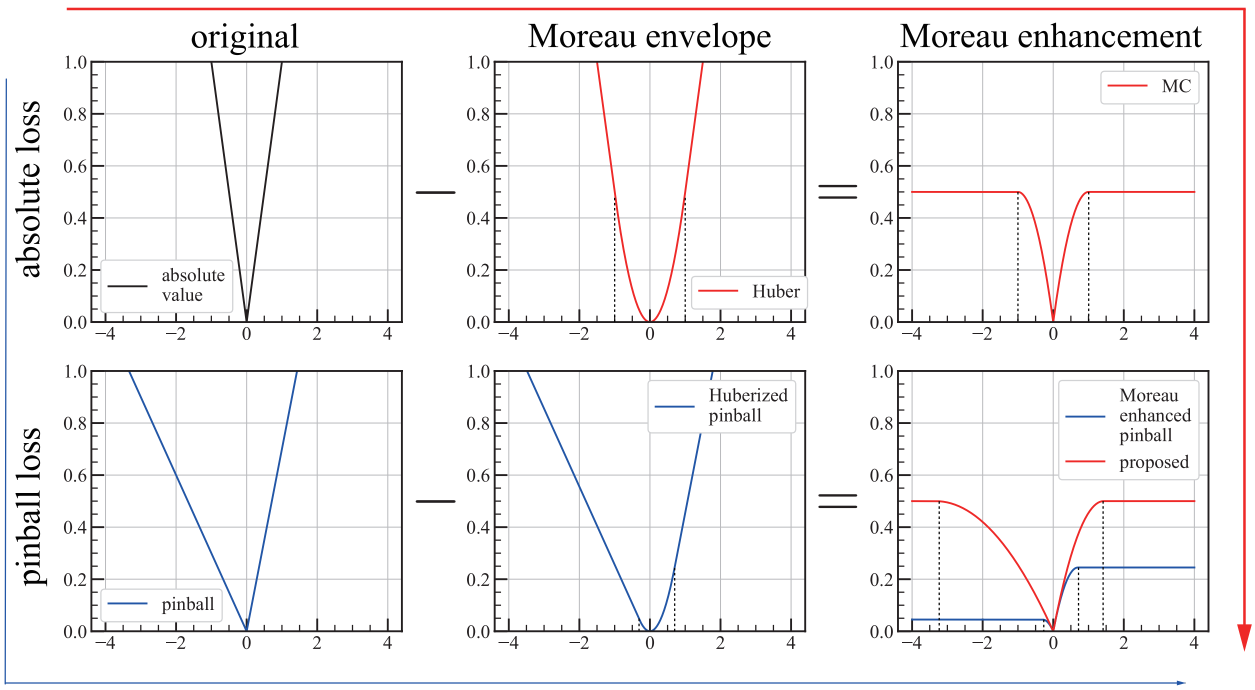


Fig. 3: The generation processes of the proposed loss $\phi_{\alpha,\gamma}^{\text{MC-pin}}$ and the Moreau-enhanced pinball loss $\phi_{\alpha,\gamma}$ for $\alpha = 0.9$.

the constants after saturation for $\phi_{\alpha,\gamma}$ significantly differ from each other on both sides of the real line. Because of this, $\phi_{\alpha,\gamma}$ encourages a larger portion of data than the desirable proportion α stay on the left side if $\alpha \approx 1$ (or on the right side if $\alpha \approx 0$).

Viewing Fig. 3 under (6), moreover, it can be seen that the ranges of nonzero gradient on the left and right sides are proportional to $1-\alpha$ and α, respectively. Indeed, the quantile estimate needs to be such that the number of data on each side is proportional to α and $1-\alpha$. If ϵ for reliable data is uniformly distributed, for instance, the nonzero-gradient ranges are desired to share the same proportion. In this specific case, the ranges for $\phi_{\alpha,\gamma}$ have the reciprocal proportion, implying that $\phi_{\alpha,\gamma}$ is undesirable to separate undesirable data from (the bulk of) reliable ones. To resolve those issues mentioned above, we propose an alternative approach in the following subsection.

B. Proposed Loss Function

We define the MC-pinball function as follows:

$$\phi_{\alpha,\gamma}^{\text{MC-pin}} : \mathbb{R} \to [0,+\infty) : z \mapsto \phi_{\gamma}^{\text{MC}} \circ \rho_\alpha(z)$$
$$= \begin{cases} \alpha z - \dfrac{\alpha^2}{2\gamma} z^2, & \text{if } 0 \le z \le \dfrac{1}{\alpha}\gamma, \\ -(1-\alpha)z - \dfrac{(1-\alpha)^2}{2\gamma} z^2, & \text{if } -\dfrac{1}{(1-\alpha)}\gamma < z \le 0, \\ \dfrac{1}{2}\gamma, & \text{otherwise,} \end{cases} \tag{7}$$

where $\gamma \in \mathbb{R}_{++}$ governs the points at which the gradient vanishes. The task of estimating the empirical quantile of reliable data is cast as the following minimization problem:

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^n} \Phi_{\alpha,\gamma}^{\text{MC-pin}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) := \sum_{i=1}^{m} \phi_{\alpha,\gamma}^{\text{MC-pin}}(y_i - \boldsymbol{x}_i^\top \boldsymbol{\theta}). \tag{8}$$

The MC-pinball function $\phi_{\alpha,\gamma}^{\text{MC-pin}}$ is depicted in Fig. 2. It approximates the pinball loss well in the vicinity of the origin, while it saturates on both sides of the real axis. The generation process of the proposed loss is indicated by the red arrow in Fig. 3. As stated already, the MC loss (top right) is generated by subtracting the Huber function (top middle: the Moreau envelope of $|\cdot|$) from $|\cdot|$ (top left). It is then composed with ρ_α to obtain the proposed loss $\phi_{\alpha,\gamma}^{\text{MC-pin}}$ (bottom right).

Comparing the two graphs in the bottom-right panel of Fig. 3 under (4), $\phi_{\alpha,\gamma}^{\text{MC-pin}}$ possesses the following remarkable properties: (i) the constants $\gamma/2$ after saturation coincides on both sides with each other, and (ii) the nonzero-gradient ranges are proportional to α and $1-\alpha$. These are in sharp contrast to the case of $\phi_{\alpha,\gamma}$, and $\phi_{\alpha,\gamma}^{\text{MC-pin}}$ is free from the issues of $\phi_{\alpha,\gamma}$ raised in Section III-A. Indeed, the range proportion is desirable when ϵ for reliable data is uniformly distributed, and the proposed loss works well when it is normally distributed, as shown by simulations in Section IV.

Thanks to properties (i) and (ii), a solution of the problem in (8) tends to give a robust estimate of the true quantile. To obtain an intuition, suppose, under property (ii), that the reliable data reside in the vicinity of the empirical quantile of the reliable data, and that the unreliable data are scattered in the saturated region (see Fig. 1). In this case, the empirical quantile is likely a minimizer of the MC-pinball loss for the reliable data in an approximate sense. Importantly, moreover, the unreliable data do not affect the minimizer, because a slight deviation from the empirical quantile of the reliable data does not change the losses at all for those unreliable data owing to property (i). Consequently, a solution of (8) would be a good approximation of the empirical quantile of the reliable data (and thus a reasonable estimate of the true quantile).

Remark 1. The proposed loss has the following asymptotic properties: (i) $\lim_{\gamma\to+\infty}\phi^{\text{MC-pin}}_{\alpha,\gamma} = \rho_\alpha$, and (ii) $\lim_{\gamma\downarrow 0}(\gamma/2)\phi^{\text{MC-pin}}_{\alpha,\gamma} = \|\cdot\|_0$. This implies that γ needs to be tuned appropriately. Fortunately, our numerical studies suggest that the proposed loss gives a good performance for γ chosen from a reasonably wide range.

IV. Numerical Examples

We conduct simulations to show that the proposed approach yields robust estimates of the true quantiles under a variety of situations. After presenting the simulation conditions, we show how the performance changes as the proportion β of unreliable data changes. We then show insensitivity of the proposed approach to the choice of the parameter γ.

A. Simulation conditions

Throughout, we consider the quantile regression for $\alpha = 0.1, 0.9$ to find the 80% interval with the top and bottom 10 percents excluded. The input matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ ($m = 2000$, $n = 11$) and the regression vector $\boldsymbol{\theta}_\star \in \mathbb{R}^n$ are i.i.d. with the uniform distribution $\mathcal{U}[0,1]$ and the standard Gaussian distribution, respectively.[1] The output vector is generated as $\boldsymbol{y} := \boldsymbol{X}\boldsymbol{\theta}_\star + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is i.i.d. with the densities f and g described below. The residuals ϵ_i of reliable data are generated from the Gaussian mixture distribution $f(\epsilon) := \frac{1}{3}\frac{1}{\sqrt{2\pi\cdot 0.1^2}}\exp\left(-\frac{(\epsilon+0.5)^2}{2\cdot 0.1^2}\right) + \frac{2}{3}\frac{1}{\sqrt{2\pi\cdot 0.2^2}}\exp\left(-\frac{(\epsilon-0.5)^2}{2\cdot 0.2^2}\right)$. For the generation of ϵ_i of unreliable data, we consider three types of distributions (see Fig. 4): (i) the uniform distribution $g_1(\epsilon) := \begin{cases}\frac{1}{2a}, & -a \le \epsilon \le a,\\ 0, & \text{otherwise,}\end{cases}$ for $a := 50$, (ii) the gamma distribution $g_2(\epsilon) := b\exp\left(-b\epsilon\right), \epsilon > 0$, for $b := 100$, and (iii) the point mass $g_3(\epsilon) := \delta(\epsilon - c)$ for $c := 100$, where δ is Dirac's delta.

Two experiments are conducted.

Experiment A: The performance is compared for different proportions β of unreliable data for the tuned parameter $\gamma := 2.0$.

Experiment B: The performance of the proposed approach is tested for different saturation factors γ for $\beta = 0.3$.

To seek for a stationary point of the proposed loss in (8), we employ Algorithm 1, where

$$\nabla_{\boldsymbol{\theta}}\Phi^{\text{MC-pin}}_{\alpha,\gamma}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) = \sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}}\phi^{\text{MC-pin}}_{\alpha,\gamma}(y_i - \boldsymbol{x}_i^\top\boldsymbol{\theta}). \quad (9)$$

[1]Following the convention, the last component of each row vector is a constant so that the last component of $\boldsymbol{\theta}_\star$ represents the bias term.

Algorithm 1 generalized gradient descent

Set $\boldsymbol{\theta}_0 := \mathbf{0}$, $\gamma \in \mathbb{R}_{++}$, $\mu \in \mathbb{R}_{++}$.
For $k = 0, 1, 2, \cdots$
$\boldsymbol{\theta}_{k+1} := \boldsymbol{\theta}_k - \mu\nabla_{\boldsymbol{\theta}}\Phi^{\text{MC-pin}}_{\alpha,\gamma}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}_k)$

Here, $\nabla_{\boldsymbol{\theta}}\phi^{\text{MC-pin}}_{\alpha,\gamma}(y_i - \boldsymbol{x}_i^\top\boldsymbol{\theta}) = -\nabla_z\phi^{\text{MC-pin}}_{\alpha,\gamma}(y_i - \boldsymbol{x}_i^\top\boldsymbol{\theta})\boldsymbol{x}_i$, where

$$\nabla_z\phi^{\text{MC-pin}}_{\alpha,\gamma}(z) \in \partial_z\phi^{\text{MC-pin}}_{\alpha,\gamma}(z) = \quad (10)$$

$$\begin{cases}\left\{\alpha - \alpha^2\frac{z}{\gamma}\right\}, & \text{if } 0 < z \le \frac{1}{\alpha}\gamma,\\ \left\{-(1-\alpha) - (1-\alpha)^2\frac{z}{\gamma}\right\}, & \text{if } -\frac{1}{1-\alpha}\gamma \le z < 0,\\ [-(1-\alpha), \alpha], & \text{if } z = 0,\\ \{0\}, & \text{otherwise.}\end{cases} \quad (11)$$

Here, ∂_z is Clarke's generalized gradient operator with respect to z [14]. In simulations, we simply let $\nabla_z\phi^{\text{MC-pin}}_{\alpha,\gamma}(0) = 0 \in [-(1-\alpha), \alpha]$.

The stepsize is set to $\mu = 0.001$. The performance is evaluated with test datasets of sample size 1000 (generated in the same way as the training datasets) by two metrics: (i) the system mismatch $\frac{\|\hat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\alpha,\star}\|_2^2}{\|\boldsymbol{\theta}_{\alpha,\star}\|_2^2}$ with the ℓ_2 norm $\|\cdot\|_2$, where $\hat{\boldsymbol{\theta}}_\alpha$ is the estimate of the true αth quantile $\boldsymbol{\theta}_{\alpha,\star}$, and (ii) coverage rate $\frac{1}{|\mathcal{I}_\text{r}|}\sum_{i\in\mathcal{I}_\text{r}} 1_{S_y(\boldsymbol{x}_i)}(y_i)$, where $|\mathcal{I}_\text{r}|$ is the number of reliable data, and

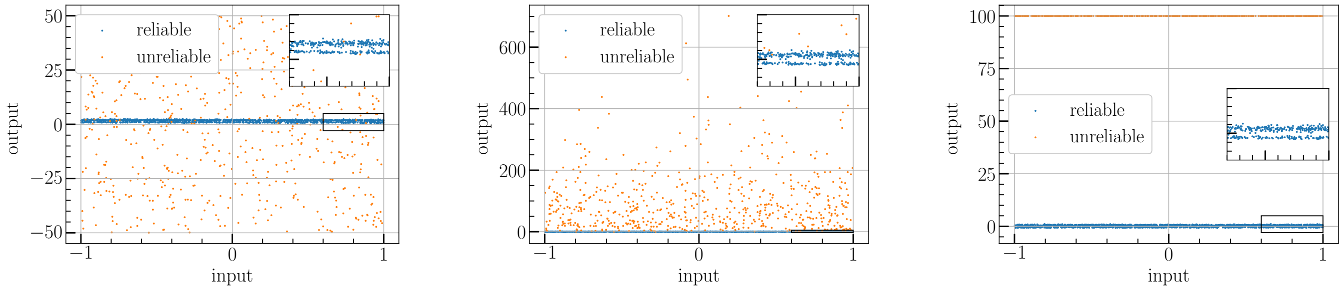
$$1_{S(\boldsymbol{x}_i)}(y_i) := \begin{cases}1 & \text{if } y_i \in S(\boldsymbol{x}_i) := [\boldsymbol{x}_i^\top\hat{\boldsymbol{\theta}}_{0.1}, \boldsymbol{x}_i^\top\hat{\boldsymbol{\theta}}_{0.9}],\\ 0 & \text{otherwise.}\end{cases} \quad (12)$$

The results are averaged over 300 independent trials.

B. Experiment A: Robustness under different proportions β of unreliable data

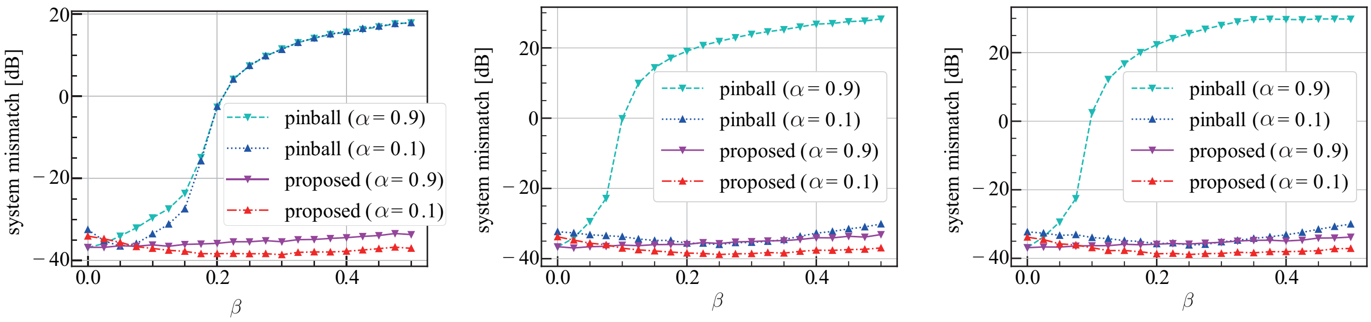
The results are plotted in Fig. 5 for each density of unreliable data. For all types of density, the proposed approach attains low system mismatches and coverage rates close to the target rate 0.8 ($= 0.9 - 0.1$) over the whole range of β. (The average error rates of coverage rate for each distribution g_i were (d) 6.0×10^{-3}, (e) 6.0×10^{-3}, and (f) 6.5×10^{-3}.) This means that the proposed approach enjoys remarkable robustness for quantile regression even when the proportion of unreliable data is large. As expected, on the other hand, the performance of the pinball loss deteriorates significantly.

To be more specific, we focus on the case of uniform distribution g_1. In this case, unreliable data arise both above and below the region of reliable data, making the upper and lower quantile estimates be larger and smaller, respectively, so that the interval expands with β. This is the reason why both upper and lower mismatches (the system mismatches of both upper and lower quantiles) increase with β. Let us now turn our attention to the cases of g_2 and g_3. In contrast to the case of g_1, unreliable data arise only above the reliable-data region.

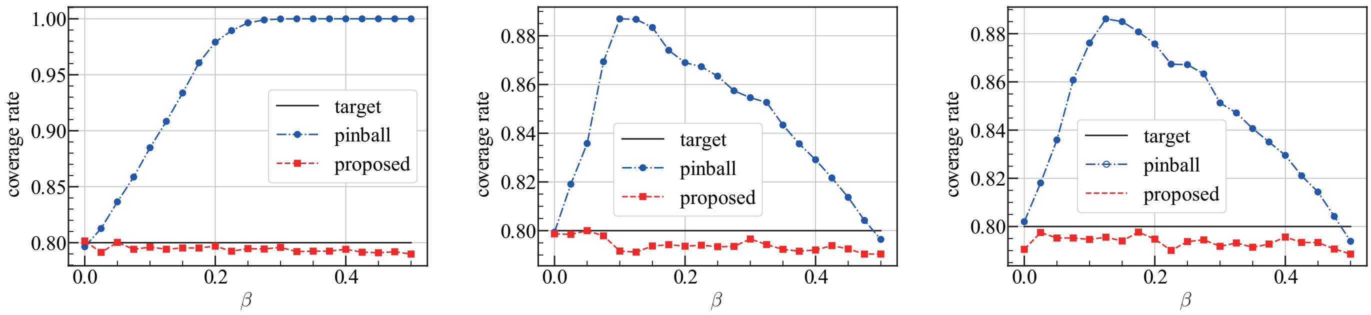


(a) uniform distribution g_1 **(b)** gamma distribution g_2 **(c)** point mass g_3

Fig. 4: Samples of ε generated from $0.7f + 0.3g_i$, $i \in \{1,2,3\}$, i.e., $\beta := 0.3$.



(a) uniform distribution g_1 **(b)** gamma distribution g_2 **(c)** point mass g_3



(d) uniform distribution g_1 **(e)** gamma distribution g_2 **(f)** point mass g_3

Fig. 5: Results of Experiment A: performance comparisons for different proportions of unreliable data.

Because of this, the upper mismatch increases sharply when β reaches 0.1 so that $y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\theta}}_{0.9}$ exceeds some part of ϵ_is of the unreliable data. Once $\boldsymbol{x}_i^\top \hat{\boldsymbol{\theta}}_{0.9}$ exceeds all y_is of the "reliable" data, the coverage rate starts to decrease due to the gradual increase of $\boldsymbol{x}_i^\top \hat{\boldsymbol{\theta}}_{0.1}$.

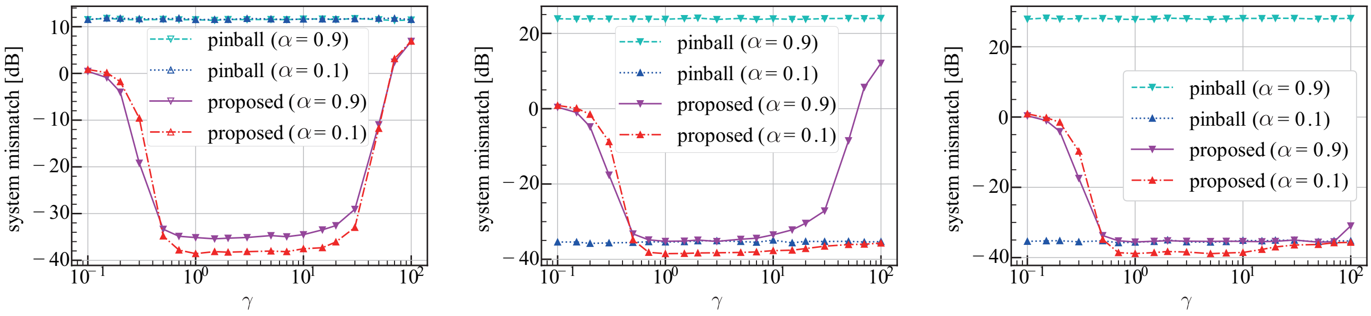
C. Experiment B: Insensitivity to the choice of γ

Fig. 6 shows the results for Experiment B. For all densities g_i, the performance is stable in the neighborhood of the best γ in both metrics. Remarkably, the range of good γ is common among the three cases; i.e., a $\gamma \in [1, 10]$ gives reasonably good performance for all g_is. This indicates the saturation parameter γ would be simple to tune. This is a potential advantage of the proposed approach. As expected from the arguments presented in Remark 1, a too large or too small γ gives deteriorated performance.

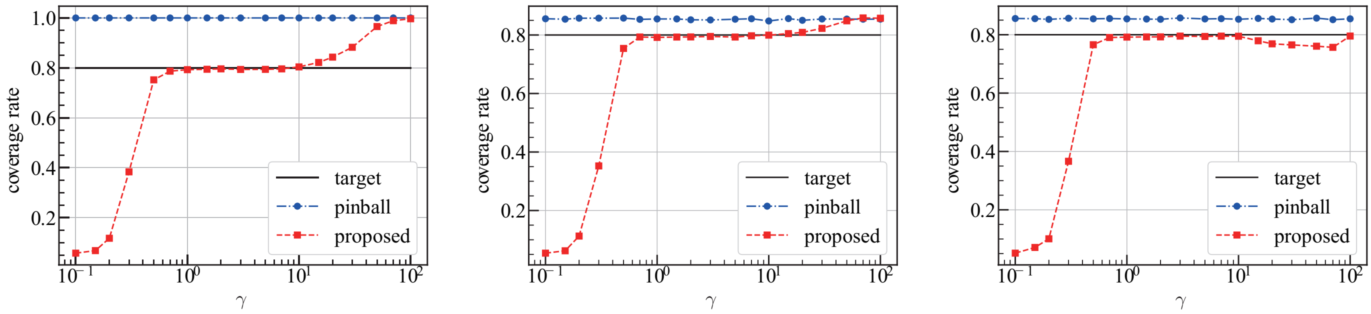
V. Concluding Remarks

We proposed the MC-pinball loss for the task of robust quantile regression in the presence of unreliable data. The proposed loss is the composition of the MC penalty and the pinball loss, enjoying the nice properties as opposed to the Moreau-enhanced pinball loss. The numerical examples suggested that the proposed approach yields robust estimates of the true quantile in the presence of unreliable data. The potential benefit of the proposed approach was also shown in terms of the parameter tuning.

Although the proposed loss is nonconvex, there was no



(a) uniform distribution g_1 **(b)** gamma distribution g_2 **(c)** point mass g_3



(d) uniform distribution g_1 **(e)** gamma distribution g_2 **(f)** point mass g_3

Fig. 6: Results of Experiment B: insensitivity to the choice of γ.

issue of dependency on the initialization nor local minima etc. in our simulations. This may suggest that the proposed loss could have benign nonconvex landscapes, which needs further detailed investigations. Sensor malfunctions or measurement errors (which could be the source of unreliable data) are typically tackled at the hardware level, and the human errors are handled by costly preprocessing. Those errors could be supported by the proposed approach at the software level, and this is a potential advantage from the practical viewpoint.

References

[1] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.

[2] C. Gallego-Castillo, R. Bessa, L. Cavalcante, and O. Lopez-Garcia, "On-line quantile regression in the RKHS (Reproducing Kernel Hilbert Space) for operational probabilistic forecasting of wind power," *Energy*, vol. 113, pp. 355–365, 2016.

[3] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, vol. 7, no. 1, pp. 47–54, 2004.

[4] S. B. Taieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2448–2455, 2016.

[5] P. S. Martins and P. T. Pereira, "Does education reduce wage inequality?: Quantile regression evidence from 16 countries," *Labour economics*, vol. 11, no. 3, pp. 355–371, 2004.

[6] G. L. Gaeta, G. Lubrano Lavadera, and F. Pastore, "Overeducation wage penalty among ph.d. holders: an unconditional quantile regression analysis on italian data," *International journal of manpower*, vol. 44, no. 6, pp. 1096–1117, 2023.

[7] F. Jareño, M. d. l. O. González, M. Tolentino, and K. Sierra, "Bitcoin and gold price returns: A quantile regression and nardl analysis," *Resources policy*, vol. 67, pp. 101 666–, 2020.

[8] T. Ichinose, M. Yukawa, and R. L. Cavalcante, "Online kernel-based quantile regression using huberized pinball loss," in *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 2023, pp. 1803–1807.

[9] Y. Romano, E. Patterson, and E. Candes, "Conformalized quantile regression," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[10] M. Yukawa, H. Kaneko, K. Suzuki, and I. Yamada, "Linearly-involved Moreau-enhanced-over-subspace model: Debiased sparse modeling and stable outlier-robust regression," *IEEE Trans. Signal Process.*, vol. 71, pp. 1232–1247, 2023.

[11] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, vol. 38, no. 2, pp. 894–942, Apr. 2010.

[12] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, Sep. 2017.

[13] J. Abe, M. Yamagishi, and I. Yamada, "Linearly involved generalized moreau enhanced models and their proximal splitting algorithm under overall convexity condition," *Inverse Problems*, vol. 36, no. 3, pp. 1–36, 2020.

[14] F. H. Clarke, "Generalized gradients and applications," *Transactions of the American Mathematical Society*, vol. 205, pp. 247–262, 1975.