Deviation Theorems in Quantile Estimation and Online Quantile Regression

Chris Junchi Li[◇]

Department of Electrical Engineering and Computer Sciences[◇]
University of California, Berkeley

August 31, 2024

Abstract

This paper presents new theoretical insights and algorithms for quantile estimation and regression in streaming data environments. We explore moderate and large deviation principles within quantile estimation, offering refined approximations that bridge the gap between central limit theorem results and large deviation principles. Additionally, we introduce an efficient Online Quantile Regression (`Online-QR`) algorithm designed for real-time data processing. By leveraging a Bayesian framework, our method reformulates the traditional quantile regression problem into a least squares problem, enhancing computational efficiency while maintaining high accuracy. We also extend the `Online-QR` to support Multiple Quantile Regression (`Multiple-QR`), enabling simultaneous estimation across multiple quantile levels. Theoretical analysis provides asymptotic properties, including unbiasedness, asymptotic normality, and linear convergence of the proposed estimators. Our findings are significant for applications requiring robust, real-time analysis of skewed or heterogeneous data streams.

Keywords: Online Quantile Regression; Moderate and Large Deviations; Bayesian Inference; Streaming Data Processing; Multiple Quantile Regression

1 Introduction

Quantile estimation and regression are robust statistical methods widely used in analyzing the relationship between a response variable and a set of covariates. Unlike traditional regression methods that focus on the mean, quantile regression provides insights into different points of the conditional distribution of the response variable. This approach is particularly useful in applications such as risk management and econometrics, where understanding the distribution of outcomes, rather than just the average, is crucial.

The estimation of quantiles from empirical data has been a focal point in statistical analysis, with extensive research dedicated to understanding the asymptotic properties of sample quantiles. These properties are often studied through the lens of moderate and large deviation principles, which provide essential insights into the likelihood of significant deviations in quantile estimates as the sample size increases. Moderate deviations bridge the gap between the central limit theorem, which addresses small fluctuations, and large deviations, which consider the probability of rare events. Understanding these deviations is critical for deriving statistical guarantees and assessing the performance of quantile-based methods under varying conditions.

In parallel, *quantile regression* (QR) is a powerful tool that does not assume a specific distribution for the error term, making it highly effective for handling skewed or heterogeneous data. However, traditional QR methods face significant challenges in dynamic environments where data arrive in streams, requiring real-time updates. Conventional QR techniques, which rely on static datasets, are often impractical for real-time applications due to their computational complexity and storage demands.

To address these limitations, we propose an *Online Quantile Regression* (`Online-QR`) algorithm designed for efficient real-time processing of streaming data. Our approach reformulates the check loss optimization problem inherent in QR as a least squares problem, significantly enhancing computational efficiency. By leveraging a Bayesian framework, the `Online-QR` algorithm performs online updates using the posterior distribution from previous data as the prior for new data. This method allows for rapid adaptation to new information, maintaining high accuracy while supporting the demands of real-time data streams.

We extend the `Online-QR` algorithm to *Multiple Quantile Regression* (`Multiple-QR`), providing a flexible tool for simultaneous estimation across multiple quantile levels. This extension is particularly valuable in applications where understanding the full distribution of a response variable is necessary.

Contributions This paper makes several key contributions. First, we offer new theoretical insights into the moderate and large deviations in quantile estimation, providing refined statistical guarantees for quantile-based methods in various contexts. Second, we introduce a novel `Online-QR` algorithm that significantly improves the efficiency of quantile regression in streaming data environments. Finally, we extend this algorithm to `Multiple-QR`, broadening its applicability.

Organization The remainder of this paper is organized as follows. Section 2 delves into new deviation results in quantile estimation. Section 3 briefly reviews Gibbs sampling for Bayesian QR and provides a detailed description of the proposed `Online-QR` algorithm. In Section 4, we extend `Online-QR` to the `Multiple-QR` model. Section 5 presents the theoretical analysis for `Online-QR`, including the unbiasedness, asymptotic normality, linear convergence of the `Online-QR` estimator, and the regret growth rate of the `Online-QR` learning procedure. All technical proofs are deferred to Section 6. Section 7 offer concluding remarks.

Notations We write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$. Let $\Phi(\cdot)$ denote the cumulative distribution function of a standard normal random variable. Specifically, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\{-t^2/2\}\, dt$ represents the standard normal distribution function. Additional notations will be introduced as they appear.

2 Moderate and large deviations in quantile estimation

This section aims to advance the theoretical understanding of sample quantiles by deriving new results within the framework of moderate and large deviations. We build on existing work in this area, particularly the Bahadur-Rao large deviations theory, to establish tighter bounds and uncover more nuanced behavior of quantile estimators. Our approach also investigates Cramér-type deviations, further enriching the analysis of empirical quantile functions.

Mathematically, recall the definitions of the quantile of population and the quantile of sample. Assume that $\{X_i, 1 \le i \le n\}$ is a sample of a population X with a common distribution function $F(x)$. For $0 < p < 1$, denote

$$x_p := F^{-1}(p) = \inf\{x : F(x) \ge p\}$$

the p-th quantile of $F(x)$. An estimator of x_p is given by the sample p-th quantile defined as follows:

$$x_{n,p} := F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\}$$

where

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x), \quad x \in \mathbf{R}$$

is the empirical distribution function of the sample $\{X_i, 1 \leq i \leq n\}$. Here, $\mathbf{1}(A)$ denotes the indicator function of set A.

There are a number of literatures to study the asymptotic properties for sample quantiles. If $x_{n,p}$ is the unique solution of the equation $F(x-) \leq p \leq F(x)$, then $x_{n,p} \to x_p$ a.e.; Assume that $F(x)$ have a continuous density function $f(x)$ in a neighborhood of x_p such that $f(x_p) > 0$. Then $\frac{\sqrt{n}f(x_p)(x_{n,p}-x_p)}{\sqrt{p(1-p)}}$ converges to the standard normal random variable in distribution, see Serfling [Ser09]. Suppose that $F(x)$ is twice differentiable at x_p, then Bahadur [Bah66] proved that

$$x_{n,p} = x_p + \frac{p - F_n(x_p)}{f(x_p)} + R_n, \quad a.e.$$

where $R_n = O(n^{-3/4}(\log n)^{3/4})$ a.e., $n \to \infty$. Xu and Miao [XM11] (see also Miao, Chen and Xu [MCX10] for order statistics) obtained the following moderate deviation principles for $x_{n,p} - x_p$. For any sequence of real numbers $\{a_n\}_{n\geq 1}$ satisfying $a_n \to \infty$ and $a_n/\sqrt{n} \to 0$ as $n \to \infty$, it holds for any $r > 0$,

$$\lim_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\bigg(\frac{\sqrt{n}}{a_n}|x_{n,p} - x_p| \geq t\bigg) = -\frac{f(x_p)^2 r^2}{2p(1-p)} \tag{1}$$

They also obtained the following large deviation principles for $x_{n,p} - x_p$: the following two equality hold for any $t > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\bigg(x_{n,p} - x_p \geq t\bigg) = -\inf_{y\geq 1-p} \Lambda_+^*(y) \tag{2}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\bigg(x_{n,p} - x_p \leq -t\bigg) = -\inf_{y\geq p} \Lambda_-^*(y) \tag{3}$$

where

$$\Lambda_+^*(y) = y \log \frac{y}{F(x_p+t)} + (1-y)\log\frac{1-y}{1-F(x_p+t)}$$

and

$$\Lambda_-^*(y) = y \log \frac{y}{F(x_p-t)} + (1-y)\log\frac{1-y}{1-F(x_p-t)}$$

In this section, we are interested in sharp moderate and large deviations between the quantiles of population and the quantiles of samples. More precisely, we establish Cramér type moderate deviations and Bahadur-Rao type large deviations for $x_{n,p} - x_p$. Our results refine the moderate and large deviation principle results (1)–(3).

Main results For brevity, denote

$$R_n(x,p) = \frac{\sqrt{n} f(x_p)(x_{n,p} - x_p)}{\sqrt{p(1-p)}}, \qquad p \in (0,1)$$

Denote $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\{-t^2/2\} dt$ the standard normal distribution function. The following theorem gives a Cramér type moderate deviation for sample quantiles.

Theorem 1. *Let $f(x)$ be the density function of X and let $p \in (0,1)$. If $f'(x)$ is bounded in a neighborhood of $x = x_p$ and $f(x_p) > 0$, then it holds*

$$\log \frac{\mathbb{P}\big(\pm R_n(x,p) \geq t \big)}{1 - \Phi(t)} = O\Big(\frac{1+t^3}{\sqrt{n}} \Big) \tag{4}$$

uniformly for $0 \leq t = o(\sqrt{n})$ as $n \to \infty$.

Remark. Let us comment on the result of Theorem 1.

(i) Assume that $f'(x)$ is uniformly bounded on $\mathbf{R}$ and that $f(x)$ is positive for all $x \in \mathbf{R}$. When p is replaced by p_n which may depend on n, by inspecting the proof of Theorem 1, the following equality

$$\log \frac{\mathbb{P}\big(\pm R_n(x,p_n) \geq t \big)}{1 - \Phi(t)} = O\Big(\frac{1+t^3}{\sqrt{np_n(1-p_n)}} \Big) \tag{5}$$

holds uniformly for $0 \leq t = o(\sqrt{np_n(1-p_n)}\,)$ as $n \to \infty$. Clearly, if $p_n(1-p_n) \to 0$ as $n \to \infty$, then the last range tends to smaller than $0 \leq t = o(\sqrt{n}\,)$.

(ii) By an argument similar to the proof of Corollary 3 in [FHM21], the following moderate deviation principle (MDP) result is a consequence of Theorem 1. Let $\{a_n\}_{n\geq 1}$ be a sequence of real numbers satisfying $a_n \to \infty$ and $a_n/\sqrt{n} \to 0$ as $n \to \infty$. Then for each Borel set B,

$$\begin{aligned} - \inf_{x \in B^o} \frac{x^2}{2} \quad &\leq \quad \liminf_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\frac{R_n(x,p)}{a_n} \in B \Big) \\ &\leq \quad \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\frac{R_n(x,p)}{a_n} \in B \Big) \leq - \inf_{x \in \overline{B}} \frac{x^2}{2} \end{aligned} \tag{6}$$

where B^o and $\overline{B}$ denote the interior and the closure of B, respectively. When B is $[t,\infty)$ or $(-\infty,t]$ for some $t > 0$, the MDP result (6) has been established by Xu and Miao [XM11] (cf. equality (1)). Notice that, in Xu and Miao [XM11], MDP result holds without the assumption that $f'(x)$ is bounded in a neighborhood of $x = x_p$.

Using the inequality $|e^x - 1| \leq e^c|x|$ valid for $|x| \leq c$, from Theorem 1, we obtain the following result about the relative errors of normal approximations.

Corollary 1. *Assume that the conditions of Theorem 1 are satisfied. Then it holds*

$$\frac{\mathbb{P}\big(\pm R_n(x,p) \geq t \big)}{1 - \Phi(t)} = 1 + O\Big(\frac{1+t^3}{\sqrt{n}} \Big) \tag{7}$$

uniformly for $0 \leq t = O(n^{1/6})$ as $n \to \infty$, which implies that

$$\frac{\mathbb{P}\big(\pm R_n(x,p) \geq t \big)}{1 - \Phi(t)} = 1 + o(1) \tag{8}$$

uniformly for $0 \leq x = o(n^{1/6})$.

From (7) in Corollary 1, by an argument similar to the proof of Corollary 2.2 in Fan *et al.* [FGLS20], we can obtain the following Berry-Esseen bound:

$$\sup_{t\in\mathbf{R}}\Big|\mathbb{P}\big(\pm R_n(x,p)\le t\big)-\Phi(t)\Big| = O\Big(\frac{1}{\sqrt{n}}\Big), \quad n\to\infty \tag{9}$$

The last convergence rate coincides with the classical result established by Reiss [Rei74].

Theorem 1 is devoted to the moderate deviations. For sharp large deviations, we have the following Bahadur-Rao type large deviation expansions.

Theorem 2. *For any $t\ge 0$, it holds*

$$\mathbb{P}\big(x_{n,p}-x_p\ge t\big)=\frac{1}{\tau_t^+\,\sigma_p\sqrt{2\pi n}}e^{-n\Lambda^+(t)}\Big[1+o(1)\Big], \quad n\to\infty$$

where

$$\tau_t^+=\log\frac{F(x_p+t)(1-p)}{p(1-F(x_p+t))}, \quad \sigma_p=\sqrt{p(1-p)} \quad \textit{and}$$

$$\Lambda^+(t)=p\log\frac{p}{F(x_p+t)}+(1-p)\log\frac{1-p}{1-F(x_p+t)}$$

Similarly, it also holds for any $t\ge 0$,

$$\mathbb{P}\big(x_{n,p}-x_p\le -t\big)=\frac{1}{\tau_t^-\,\sigma_p\sqrt{2\pi n}}e^{-n\Lambda^-(t)}\Big[1+o(1)\Big], \quad n\to\infty$$

where

$$\tau_t^-=\log\frac{p(1-F(x_p-t))}{F(x_p-t)(1-p)} \quad \textit{and}$$

$$\Lambda^-(t)=p\log\frac{p}{F(x_p-t)}+(1-p)\log\frac{1-p}{1-F(x_p-t)}$$

Denote c a finite and positive constant which does not depend on n. For two sequences of positive numbers $(a_n)_{n\ge1}$ and $(b_n)_{n\ge1}$, write $a_n\asymp b_n$ if there exists a c such that $a_n/c\le b_n\le c\,a_n$ for all sufficiently large n. By Theorem 2, we have for any given constants $t>0$ and $p\in(0,1)$,

$$\mathbb{P}\big(x_{n,p}-x_p\ge t\big)\asymp\frac{1}{\sqrt{n}}e^{-n\Lambda^+(t)} \quad \text{and} \quad \mathbb{P}\big(x_{n,p}-x_p\le -t\big)\asymp\frac{1}{\sqrt{n}}e^{-n\Lambda^-(t)}, \quad n\to\infty$$

In particular, from the last line, we recover the following large deviation principle (LDP) result of Xu and Miao [XM11]: for any $t>0$, the following equalities hold

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(x_{n,p}-x_p\ge t\big)=-\Lambda^+(t) \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(x_{n,p}-x_p\le -t\big)=-\Lambda^-(t)$$

see equalities (2)–(3). Notice that $\Lambda^+(t)=\inf_{y\ge1-p}\Lambda_+^*(y)$ and $\Lambda^-(t)=\inf_{y\ge p}\Lambda_-^*(y)$, see Remark 1 in Xu and Miao [XM11].

3 The `Online-QR` algorithm

Recall that linear QR assumes that the τ th $(0 < \tau < 1)$ conditional quantile of y given x is

$$Q_{y|x}(\tau) = \boldsymbol{\beta}^o(\tau) + x^\top \boldsymbol{\beta}(\tau)$$

Given a data set $\{(\boldsymbol{x}_i, y_i), i = 1, 2, \dots, N\}$, the $\mathbb{R}^{p+1}$-valued unknown regression parameters $\boldsymbol{\beta}^{\mathrm{QR}}(\tau) = (\boldsymbol{\beta}^o(\tau), \boldsymbol{\beta}(\tau))^\top$ at a fixed quantile level is estimated by

$$\widehat{\boldsymbol{\beta}}^{\mathrm{QR}}(\tau) = \arg\min_{\boldsymbol{\beta}^o, \boldsymbol{\beta}} \sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{\beta}^o - x_i^\top \boldsymbol{\beta}) \tag{10}$$

where $\rho_\tau(u) = u[\tau - \mathbb{I}(u < 0)]$ $(u \in \mathbb{R})$ is the check loss function and $\mathbb{I}(\cdot)$ is the indicator function. QR [KBJ78] is a powerful tool for studying the dependency of a response variable $y \in \mathbb{R}$ on a set of covariates $\boldsymbol{x} = (x_1, x_2, \dots, x_p)^T \in \mathbb{R}^p$. Through modeling the conditional quantiles of y given x, it gives a comprehensive portrayal for their relationship. Moreover, QR does not impose distributional assumptions on the error and is hence suitable for analyzing skewed or heterogeneous data. Its applications appear widely in economics [CH06, FFL09, FKM13, FKMM22] and finance [MP08, AGPS09]. For example, studying the herding behavior in stock markets of Finland [Saa08] and India [BK22], analyzing the inflation-return [AP12] and volatility-return relationships [Bad13], and discussing the impacts of economic development and schooling on wage structure [Buc98, ACFV06].

Traditional statistical analysis is often based on a static data set. That is, the analysis is performed after all data become available. However, stream data, as data blocks arriving in an online fashion in time, demand frequent update of the analysis result as more and more data are collected [ZW07, GSR13]. Due to rapid development of information technology, stream data have become ubiquitous in modern real-time data processing infrastructures. For example, financial institutions track the changes of stock price in real-time and calculate the value-at-risk to re-balance portfolio [WWL22]. Land and resource management departments record the real-time house prices to timely grasp market dynamics [EI97, GN17]. E-commerce companies analyze users' browsing and purchasing history in real-time for more accurate personalized recommendations [Ama13, FLL23]. Although collected over time, compared to traditional time series, the analysis for stream data usually does not focus on modeling the temporal correlation structure [LKLC03]. Rather, the two key issues are (10) the storage challenge that old data are already discarded when new data arrive, and (11) the speed challenge that the analysis must be updated in a very fast way to support real-time information processing. Before the old data are discarded, information therein is typically compressed efficiently as a fixed dimensional object whose size does not grow over time. In this paper, we assume that at a given time point t, only the current data block $(X_t, \boldsymbol{y}_t)$ and the compressed information based on the previous $(t-1)$ data blocks are available (see Fig. 1), and term the data processing and information aggregation at a single data block as local processing.

Conventionally, because of the non-smoothness in the check loss function, the QR estimation problem (10) cannot be directly solved by gradient-based methods. Rather it is written as a linear programming (LP) problem and solved by the interior point (IP) method [PK97] whose complexity is $O\left(N^{1+\alpha} p^3 \log N\right)$, $\alpha \in (0, 1/2)$ [HPTZ23]. Its computation under the big data context has undergone rapid development recently. Common strategies include subsampling and divide-and-conquer. The subsampling methods rely on analyzing a subset of the big data. Other than the naive simple random sampling, the subset can be chosen by projecting the big data to a lower dimensional space [YMM14] or based on optimal design principles [AWYZ21]. The divide-and-conquer technique processes partitions of the entire data in parallel and an aggregated estimator is then computed to approximate the QR estimator $\widehat{\boldsymbol{\beta}}_{QR}(\tau)$. According to the number of rounds for information aggregation required by the estimation process, the divide-and-conquer algorithms can be classified into one-shot algorithms [CLZ19, VCC19, CLMY20, CZ20] and multi-shot algorithms [YL17, YLW17, FLL23, WYW$^+$23, LLSS24]. However, most of these techniques are only applicable to static big data but not stream data except the one-shot divide-and-conquer algorithms when viewing stream data as data blocks partitioned over the time domain. The subsampling methods obviously require the availability of the entire data set to sample from it, and the multi-shot divide-and-conquer algorithms need to repeatedly access the partitioned data blocks, which is infeasible in stream data. On the other hand, while the one-shot divideand-conquer algorithms can be naturally extended to stream data by implementing the aggregation in a block-by-block manner, their asymptotic theory often requires the block size to grow to infinity, while the block size in stream data is generally determined by hardware capacity and hence fixed. In addition, updates based on direct applications of one-shot divide-and-conquer algorithms may not be efficient enough to support the frequent update in the stream data context.

Another set of techniques closely related to our context is the recently proposed renewable regression estimators, including linear regression [vWB22], generalized linear model [LS20] and linear mixed effect model [LS23]. Direct applications of the renewable estimator to QR are straightforward extensions of the one-shot divide-and-conquer technique, such as the simple average estimation (SAE) method developed in [ZDW13] and the renewable QR estimation method proposed in [WWL22]. However, local processing in both methods require solving a QR problem, resulting in updates that are too slow for real-time processing of stream data. For example, the electronic health record data collected by the Scientific Registry of Transplant Recipients are updated every 10 min [LS23]; the partial discharges signal data in power grids are collected by ultra-high frequency sensors every 0.4 ns [WWSH17].

Our proposed online QR (`Online-QR`) algorithm was partly motivated by Bayesian QR for that online updates occur naturally under the Bayesian setup, i.e., using the posterior distribution obtained from the past analysis as the prior information for later analysis. Bayesian QR uses a linear model with an asymmetric Laplace (AL) error term to achieve equivalence between the original QR estimator and the Bayesian maximum-a-posteriori (MAP) estimator [YM01]. However, posterior simulation in Bayesian inference does not meet the computational speed requirement for stream data. In each update of the `Online-QR` algorithm, we essentially impose a Gaussian prior for the regression parameter centered at the previous estimate, which is equivalent to putting a ridge penalty. Then using the scale-mixture representation of the AL distribution [KK11], we can derive the updated regression parameter estimate from a normal linear model based on a proper estimate of the scale parameter. Hence, each update in `Online-QR` only requires solving a least squares problem and significantly improves the computational efficiency. The scale parameter update is done similar to the stochastic approximation (SA) for generalized linear mixed models (GLMM) [Cha08].

1. Store $(X_t, \boldsymbol{y}_t)$ and $\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}$
2. Compute $\widetilde{\boldsymbol{v}}_t$ and $\widetilde{\boldsymbol{\beta}}_t^{\mathrm{QR}}$
3. Update $\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}$

The general idea of `Online-QR` can easily extend to other variants of QR. In addition to the standard QR, we also exemplify its application to multiple, or composite, quantile regression

(`Multiple-QR`) [ZY08].

To speed up local processing to accommodate the frequent update in stream data, we convert the check loss optimization based on each data block to a least squares problem, which is partly motivated by the Gibbs sampling in Bayesian QR. To ensure that this paper is self-contained, in this section we first briefly review the Gibbs sampler for Bayesian QR and then detail the proposed `Online-QR`. Throughout this section, we consider QR at a fixed quantile level, so we omit " τ " in all notations for simplicity.

3.1 The Gibbs sampler for Bayesian QR

The linear QR model at quantile level τ can also be written as $y = x^{*\top}\boldsymbol{\beta}^{\mathrm{QR}} + \epsilon$, where $x^* = \left(1, x^\top\right)^\top \in \mathbb{R}^{p+1}$ and $Q_\epsilon(\tau) = 0$. The most popular way to formulate Bayesian QR is proposed by [YM01], which imposed an asymmetric Laplace (AL) distribution, $\mathrm{AL}(0, 1, \tau)$, on ϵ with location parameter 0, scale parameter 1, and skew parameter τ [YZ05]. This gives the error density $f(\epsilon) = \tau(1-\tau)\exp\left\{-\rho_\tau(\epsilon)\right\}$, and the model likelihood is then

$$f\left(\boldsymbol{y} \mid \boldsymbol{\beta}^{\mathrm{QR}}\right) = \tau^N(1-\tau)^N \exp\left\{-\sum_{i=1}^{N}\rho_\tau(y_i - x_i^{*\top}\boldsymbol{\beta}^{\mathrm{QR}})\right\} \tag{11}$$

where $\boldsymbol{y} = \left(y_1, y_2, \ldots, y_N\right)^\top \in \mathbb{R}^N$. Under a flat prior, finding the Bayesian MAP estimator is equivalent to solving the original QR problem (10). Posterior simulation can then be implemented through a Gibbs sampler [KK11] based on a scale-mixture representation of the AL distribution, which is given next in Lemma 1.

Lemma 1 (Proposition 1 of [KK11])**.** *If a random variable ϵ follows $\mathrm{AL}(0, 1, \tau)$, then it can be represented as*

$$\epsilon = \xi_1 v + \xi_2\sqrt{v}u \qquad \text{where } \xi_1 = \frac{1-2\tau}{\tau(1-\tau)} \text{ and } \xi_2 = \sqrt{\frac{2}{\tau(1-\tau)}}$$

and v is a standard exponential variable and u is an (independent) standard normal variable.

Using Lemma 1, we have

$$y_i = x_i^{*\top}\boldsymbol{\beta}^{\mathrm{QR}} + \xi_1 v_i + \xi_2\sqrt{v_i}u_i \qquad i = 1, 2, \ldots, N \tag{12}$$

Conditioning on the scale parameter v_i, (12) is then a normal linear model,

$$y_i \mid \boldsymbol{\beta}^{\mathrm{QR}}, v_i \sim N(x_i^{*\top}\boldsymbol{\beta}^{\mathrm{QR}} + \xi_1 v_i, \xi_2^2 v_i) \tag{13}$$

Therefore, it is easy to derive the full conditional distributions of $\boldsymbol{\beta}^{\mathrm{QR}}$ and v_i as

$$\boldsymbol{\beta}^{\mathrm{QR}} \mid \boldsymbol{y}, \boldsymbol{v} \sim N(\widehat{\boldsymbol{\beta}}, \widehat{B}) \tag{14}$$

and

$$v_i \mid y_i, \boldsymbol{\beta}^{\mathrm{QR}} \sim \mathcal{GIG}\left(\frac{1}{2}, \widehat{\delta}_i, \widehat{\gamma}\right) \tag{15}$$

under the normal prior $\boldsymbol{\beta}^{\mathrm{QR}} \sim N(\boldsymbol{\beta}^o, B_0)$, where

$$\boldsymbol{v} = \left(v_1, v_2, \ldots, v_N\right)^\top \qquad \widehat{\boldsymbol{\beta}} = \widehat{B}\left[\sum_{i=1}^{N}\frac{x_i^*\left(y_i - \xi_1 v_i\right)}{\xi_2^2 v_i} + B_0^{-1}\boldsymbol{\beta}^o\right]$$

$$\widehat{B}^{-1}=\sum_{i=1}^{N}\frac{x_i^* x_i^{*\top}}{\xi_2^2 v_i}+B_0^{-1} \qquad \widehat{\delta}_i^2=\frac{\left(y_i-x_i^{*\top}\boldsymbol{\beta}^{\mathrm{QR}}\right)^2}{\xi_2^2} \qquad \widehat{\gamma}^2=2+\frac{\xi_1^2}{\xi_2^2}$$

Here $\mathcal{GI}(\cdot)$ represents the generalized inverse Gaussian distribution [Dag89]. Eqs. (14) and (15) then form the Gibbs sampler for Bayesian QR.

3.2 The `Online-QR` algorithm

In this subsection, we give the algorithmic details of `Online-QR`. Assume that the data block arriving at time point $t(t=1,2,\ldots)$ is $(X_t,\boldsymbol{y}_t)$, where $X_t=(x_{1t},x_{2t},\ldots,x_{n_t t})^\top\in\mathbb{R}^{n_t x_p}$ and $\boldsymbol{y}_t=(y_{1t},y_{2t},\ldots,y_{n_t t})^\top\in\mathbb{R}^{n_t}$. In the local processing at time point t, existing methods for QR estimation in stream data, including SAE [ZDW13] and the renewable QR [WWL22], compute the following local QR estimator,

$$\widehat{\boldsymbol{\beta}}_t^{\mathrm{QR}}=\arg\min_{\boldsymbol{\beta}^{\mathrm{QR}}}\sum_{i=1}^{n_t}\rho_\tau(y_{it}-x_{it}^{*\top}\boldsymbol{\beta}^{\mathrm{QR}}) \tag{16}$$

which is a LP problem with time complexity $O(n_t^{1+\alpha}p^3\log n_t)$, $\alpha\in(0,1/2)$ [HPTZ23]. Our `Online-QR` algorithm instead only solves a least squares problem and hence is significantly faster than these existing solutions.

The key observation is that, when conditioning on v_i, (12) is a normal linear model. Thus, if we can find a proper estimate $\widetilde{v}_{it}$ for v_{it} (here v_{it} is the scale parameter corresponding to (x_{it},y_{it})) based on the compressed information till time point $(t-1)$, we can construct a pseudo response $y_{it}^*=y_{it}-\xi_1\widetilde{v}_{it}$, and a local estimate of $\boldsymbol{\beta}^{\mathrm{QR}}$ can be given by the least squares estimator for model (12), i.e.,

$$\widetilde{\boldsymbol{\beta}}_t^{\mathrm{QR}}=(X_t^{*\top}X_t^*)^{-1}X_t^*\boldsymbol{y}_t^* \tag{17}$$

where $X_t^*=[\mathbf{1}_{n_t},X_t]\in\mathbb{R}^{n_t\times(p+1)}$ and $\boldsymbol{y}_t^*=(y_{1t}^*,y_{2t}^*,\ldots,y_{n_t t}^*)^\top\in\mathbb{R}^{n_t}$.

Let $\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}$ denote the `Online-QR` estimator at time point $(t-1)$ (whose estimation will be discussed later in this section). By the full conditional distribution (15) in the Gibbs sampler for Bayesian QR, we have

$$f(v_{it}\mid y_{it},\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}})\propto v_{it}^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\widehat{\delta}_{it}^2 v_{it}^{-1}+\widehat{\gamma}^2 v_{it})\right\}$$

where

$$\widehat{\delta}_{it}^2=\frac{(y_{it}-x_{it}^{*\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}})^2}{\xi_2^2}$$

We next define an "estimate" of v_{it} as

$$\widetilde{v}_{it}=\frac{\sqrt{\widehat{\delta}_{it}^2\widehat{\gamma}^2}-\frac{1-2\tau+2\tau^2}{\tau(1-\tau)}|\widehat{\gamma}|\xi_2^{-1}+\widehat{\gamma}^2}{\widehat{\gamma}^2} \tag{18}$$

This choice of $\widetilde{v}_{it}$ is developed based on the form of the intuitive "estimate", the expectation of the conditional distribution of $v_{it}\mid y_{it},\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}$, to guarantee the unbiasedness of the `Online-QR` estimator (see (32) in the proof of Theorem 3 in Section 6). This construction is inspired by the SA method developed for GLMM [Cha08], as v_i is the random intercept when viewing model (12) as a linear mixed effect model.

Algorithm 1 The `Online-QR` algorithm for QR estimation in stream data

1: **Input:** data blocks (X_t, y_t), $t = 1, 2, \dots, T$ quantile level τ, and the initial estimate $\overline{\boldsymbol{\beta}}_0^{\mathrm{QR}}$
2: Set $\xi_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $\xi_2 = \sqrt{\frac{2}{\tau(1-\tau)}}$ and $\widehat{\gamma}^2 = 2 + \frac{\xi_1^2}{\xi_2^2}$
3: **for** $t = 1, 2, \dots, T$ **do**
4: Compute $\widehat{\delta}_{it}^2 = \frac{\left(y_{it} - x_{it}^{*\top} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right)^2}{\xi_2^2}$ and $\widetilde{v}_{it} = \frac{\sqrt{\widehat{\delta}_{it}^2 \widehat{\gamma}^2} - \frac{1-2\tau+2\tau^2}{\tau(1-\tau)} |\widehat{\gamma}| \xi_2^{-1} + \widehat{\gamma}^2}{\widehat{\gamma}^2}$, $i = 1, 2, \dots, n_t$
5: Set $\boldsymbol{y}_t^* = \boldsymbol{y}_t - \xi_1 \widetilde{\boldsymbol{v}}_t$ and compute $\widetilde{\boldsymbol{\beta}}_t^{\mathrm{QR}} = \left(X_t^{*\top} X_t^*\right)^{-1} X_t^* \boldsymbol{y}_t^*$
6: Update $\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} = \frac{n_t}{\sum_{l=1}^t n_l} \widetilde{\boldsymbol{\beta}}_t^{\mathrm{QR}} + \frac{\sum_{l=1}^{t-1} n_l}{\sum_{l=1}^t n_l} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}$
7: $t = t + 1$
8: **end for**
9: return the final estimate $\overline{\boldsymbol{\beta}}_T^{\mathrm{QR}}$

Next, we discuss how to update the `Online-QR` estimator at time point t. In renewable QR, the update is done by taking a weighted average over the local estimates $\widehat{\boldsymbol{\beta}}_l^{\mathrm{QR}} (l = 1, 2, \dots, t)$. We use a different idea motivated by the online update in Bayesian inference and obtain an even simpler one. Specifically, we conduct the update by solving the following nonzero centered ridge regression problem,

$$\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} = \arg\min_{\boldsymbol{\beta}^{\mathrm{QR}}} \ \left\| y_t^* - X_t^* \boldsymbol{\beta}^{\mathrm{QR}} \right\|_2^2 + \lambda_t \left\| C_t (\boldsymbol{\beta}^{\mathrm{QR}} - \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}) \right\|_2^2 \tag{19}$$

where C_t is a given matrix. The penalty term can be viewed equivalently as setting a Gaussian prior $N(\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}, \lambda_t^{-1} (C_t^\top C_t)^{-1})$ for the Bayesian linear regression $y_t^* \mid \boldsymbol{\beta}^{\mathrm{QR}} \sim N(X_t^* \boldsymbol{\beta}^{\mathrm{QR}}, I_{n_t})$, where I_{n_t} is the n_t-dimensional identity matrix. Hence the penalty parameter λ_t serves a similar role as scaling the prior variance, i.e., how strongly we believe the current estimate should be similar to the previous one. Recently, [vWB22] used this ridge regression idea in online linear regression for big data, in which λ_t was selected by cross validation and C_t was set as I_{p+1}. Since cross validation is infeasible in stream data, we set $\lambda_t = \frac{\sum_{l=1}^{t-1} n_l}{n_t}$ from its intuitive connection to the prior variance and let $C_t = X_t^*$ to guarantee that the problem (19) has a closed-form solution given as follows.

$$\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} = \frac{n_t}{\sum_{l=1}^t n_l} \widetilde{\boldsymbol{\beta}}_t^{\mathrm{QR}} + \frac{\sum_{l=1}^{t-1} n_l}{\sum_{l=1}^t n_l} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}} \tag{20}$$

The update of our proposed `Online-QR` algorithm at time point t involves evaluating (17), (18) and (20). And we summarize `Online-QR` in Algorithm 1 and Fig. 2, where $\widetilde{\boldsymbol{v}}_t = (\widetilde{v}_{1t}, \widetilde{v}_{2t}, \dots, \widetilde{v}_{n_t t})^\top \in \mathbb{R}^{n_t}$, and $\overline{\boldsymbol{\beta}}_0^{\mathrm{QR}}$ is the initial value of the algorithm, which can be set as the local estimate based on the first data block. When the duration of the data stream is T, we denote $\overline{\boldsymbol{\beta}}_T^{\mathrm{QR}}$ as the final estimate for $\boldsymbol{\beta}^{\mathrm{QR}}$.

3.3 The space complexity of `Online-QR`

In stream data analysis, space complexity, i.e., the volume of objects compressed from historical data that are required in the algorithm update, is an important consideration [CMVW16]. It is closely related to the efficient use of storage resources, as we need to store these objects with the new data even after the old data are discarded. See Fig. 1. In our `Online-QR` algorithm, it only needs to store the fixed dimensional vector $\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}} \in R^{p+1}$ in updating $\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}$, and thus the complexity is $O(p)$. The SAE method has the same storage requirement, because it also only needs the estimate from the previous time point. In contrast, the complexity of renewable QR is higher at $O\left(p^2\right)$ since it also needs to store the matrix $\sum_{l=1}^{t-1} X_l^{*\top} X_l^*$ for constructing the weights used in averaging $\widehat{\boldsymbol{\beta}}_l^{\mathrm{QR}}$ $(l=1,2,\ldots,t)$.

4 The extension of `Online-QR` to `Multiple-QR`

The general idea of `Online-QR` developed in Section 3 can also extend to other variants of QR. In this section, we give the implementation of `Online-QR` for `Multiple-QR`.

`Multiple-QR` considers simultaneous modeling at many quantile levels $0<\tau_1<\tau_2<\ldots<\tau_S<1$ to improve the estimation efficiency of QR. It assumes identical slope parameter $\boldsymbol{\beta} \in \mathbb{R}^p$ at different quantile levels but allows the intercept $\boldsymbol{\beta}^{o\tau_s} \in \mathbb{R}(s=1,2,\ldots,S)$ to vary. The local `Multiple-QR` estimator at time point t based on $(X_t, \boldsymbol{y}_t)$ is given by

$$\left(\widehat{\boldsymbol{\beta}}_{0\tau_1 t}, \widehat{\boldsymbol{\beta}}_{0\tau_2 t}, \ldots, \widehat{\boldsymbol{\beta}}_{0\tau_S t}, \widehat{\boldsymbol{\beta}}_t\right)=\underset{\{\boldsymbol{\beta}^{o\tau_s}\}_S, \boldsymbol{\beta}}{\arg\min} \sum_{s=1}^{S} \sum_{i=1}^{n_t} \rho_{\tau_s}(y_{it}-\boldsymbol{\beta}^{o\tau_s}-x_{it}^{\top}\boldsymbol{\beta}) \tag{21}$$

For simplicity, we denote $(\boldsymbol{\beta}^{o\tau_1}, \boldsymbol{\beta}^{o\tau_2}, \ldots, \boldsymbol{\beta}^{o\tau_S}, \boldsymbol{\beta})$ and $(\widehat{\boldsymbol{\beta}}_{0\tau_1 t}, \widehat{\boldsymbol{\beta}}_{0\tau_2 t}, \ldots, \widehat{\boldsymbol{\beta}}_{0\tau_S t}, \widehat{\boldsymbol{\beta}}_t)$ as $\boldsymbol{\beta}_{\texttt{Multiple-QR}} \in \mathbb{R}^{p+S}$ and $\widehat{\boldsymbol{\beta}}_{\texttt{Multiple-QR}t} \in \mathbb{R}^{p+S}$, respectively. Note that solving (21) is even slower than solving (16) in QR as both its dimensionality (i.e., $(p+S)$) and size (i.e., $S\times n_t$) are larger. In the following, we show how to convert this problem to a least squares problem using a similar idea as in Section 3 .

The scale-mixture representation for Bayesian `Multiple-QR` [Alh16] is

$$y_{it}=\boldsymbol{\beta}^{o\tau_s}+x_{it}^{\top}\boldsymbol{\beta}+\xi_{1\tau_s} v_{it\tau_s}+\xi_{2\tau_s}\sqrt{v_{it\tau_s}}\, u_{it\tau_s} \qquad s=1,2,\ldots,S$$

where $v_{it\tau_s}$'s are standard exponential variables, $u_{it\tau_s}$'s are standard normal variables, $\xi_{1\tau_s}=\frac{1-2\tau_s}{\tau_s(1-\tau_s)}$ and $\xi_{2\tau_s}=\sqrt{\frac{2}{\tau_s(1-\tau_s)}}$. Thus, given a reasonable estimate $\widetilde{v}_{\mathrm{itt}}$ for $v_{it\tau_s}$, (21) can also be converted to the least squares problem,

$$\widetilde{\boldsymbol{\beta}}_{\texttt{Multiple-QR}t}=\underset{\{\boldsymbol{\beta}^{o\tau_s}\}_S, \boldsymbol{\beta}}{\arg\min} \frac{1}{2}\sum_{s=1}^{S}\left\|\boldsymbol{y}_{\tau_s}^{*}-\boldsymbol{\beta}^{o\tau_s}\mathbf{1}_{n_t}-X_t\boldsymbol{\beta}\right\|_2^2 \tag{22}$$

where $y_{t\tau_s}^{*}=\left(y_{1t\tau_s}^{*}, y_{2t\tau_s}^{*}, \ldots, y_{n_t t\tau_s}^{*}\right)^{\top} \in \mathbb{R}^{n_t}$ and $y_{it\tau_s}^{*}=y_{it}-\xi_{1\tau_s}\widetilde{v}_{it\tau_s}$. Following the method in Section 3, we obtain $\widetilde{v}_{it\tau_s}$ by

$$\widetilde{v}_{it\tau_s}=\frac{\sqrt{\widehat{\delta}_{it\tau_s}^{2}\widehat{\gamma}_{\tau_s}^{2}}-\frac{1-2\tau_s+2\tau_s^2}{\tau_s(1-\tau_s)}\left|\widehat{\gamma}_{\tau_s}\right|\xi_{2\tau_s}^{-1}+\widehat{\gamma}_{\tau_s}^{2}}{\widehat{\gamma}_{\tau_s}^{2}} \tag{23}$$

where $\widehat{\delta}_{it\tau_s}^{2}=\frac{\left(y_{it}-\overline{\boldsymbol{\beta}}_{0\tau_s(t-1)}-x_{it}^{\top}\overline{\boldsymbol{\beta}}_{t-1}\right)^2}{\xi_{2\tau_s}^{2}}$ and $\widehat{\gamma}_{\tau_s}^{2}=2+\frac{\xi_{1\tau_s}^{2}}{\xi_{2\tau_s}^{2}}$. And the online update of the `Online-QR` estimator of $\boldsymbol{\beta}_{\texttt{Multiple-QR}}$ at time point t is then

$$\overline{\boldsymbol{\beta}}_{\texttt{Multiple-QR}t}=\frac{n_t}{\sum_{l=1}^{t} n_l}\widetilde{\boldsymbol{\beta}}_{\texttt{Multiple-QR}t}+\frac{\sum_{l=1}^{t-1} n_l}{\sum_{l=1}^{t} n_l}\overline{\boldsymbol{\beta}}_{\texttt{Multiple-QR}(t-1)} \tag{24}$$

where $\overline{\boldsymbol{\beta}}_{\texttt{Multiple-QR}t}=(\overline{\boldsymbol{\beta}}_{0\tau_1 t}, \overline{\boldsymbol{\beta}}_{0\tau_2 t}, \ldots, \overline{\boldsymbol{\beta}}_{0\tau_S t}, \overline{\boldsymbol{\beta}}_t^{\top})^{\top}$. Eqs. (22), (23), (24) then form the `Online-QR` algorithm for `Multiple-QR` estimation.

5 Theoretical analysis

In this section, we analyze the theoretical properties of `Online-QR`. Under mild conditions, we show that $\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}$ can achieve unbiasedness after finite time and further as $t \to \infty$, this estimator is asymptotically normal. The asymptotic covariance matrix can be updated through a recursive formula. Besides, we also analyze the convergence rate of $\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}$ and the regret growth rate of the `Online-QR` learning procedure. We provide these properties in Theorems 3, 4, 5.

Theorem 3. *Assume that there exists an infinite sequence of data blocks $\{(X_t, \boldsymbol{y}_t), t = 1, 2, \ldots\}$ and for any t, $(X_t, \boldsymbol{y}_t)$ satisfies:*

(i) $\operatorname{rank}(X_t) = p$;

(ii) $y_t = (y_{1t}, y_{2t}, \ldots, y_{n_t t})^\top$ *are independent.*

Then under model (12)*, the* ***Online-QR*** *estimator can achieve unbiasedness after finite time, i.e., there exists $t_0 > 0$, such that for $t > t_0$, we have*

$$\mathbb{E}(\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}) = \boldsymbol{\beta}^{\mathrm{QR}}$$

Further, if we assume

(iii) $\lambda_{\min} \leq \lambda_j \left(\frac{X_t^\top X_t}{n_t}\right) \leq \lambda_{\max}$, *where $\lambda_{\min}$ and $\lambda_{\max}$ are two positive constants, and $\lambda_j(\cdot)$ represents the jth eigenvalue of a matrix, and*

(iv) all data block sizes are on the same magnitude, i.e., $\max_t n_t / \min_t n_t = O(1)$.

Then as $t \to \infty$, the ***Online-QR*** *estimator is asymptotically normal*

$$\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} \xrightarrow{d} N\left(\boldsymbol{\beta}^{\mathrm{QR}}, \Sigma_t\right)$$

with the asymptotic covariance matrix updated recursively by

$$\Sigma_t = \frac{\left(\sum_{l=1}^{t-1} n_l\right)^2 + 2\left(\sum_{l=1}^{t-1} n_l\right) n_t (1-2\tau)^4 + n_t^2 (1-2\tau)^2}{\left(\sum_{l=1}^{t} n_l\right)^2} \Sigma_{t-1} + \frac{n_t^2 g(\tau)}{\left(\Sigma_{l=1}^{t} n_l\right)^2} (X_t^{*\top} X_t^*)^{-1}$$

where

$$g(\tau) = \frac{1 - 2\tau + 2\tau^2 + (1-2\tau)^2\left(1 - 2\tau - 2\tau^2 + 8\tau^3 - 4\tau^4\right) - 2(1-2\tau)^4}{\tau^2(1-\tau)^2}$$

is a function of the quantile level τ.

In proving the unbiasedness of the `Online-QR` estimator, we utilize the fact that the sequence $\{\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}\}, t = 1, 2, \ldots$ is a Markov process, and the unbiasedness then comes as a result of the stationarity of this Markov process. The asymptotic normality of `Online-QR`, on the other hand, follows from the martingale central limit theorem after treating $(\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} - \boldsymbol{\beta}^{\mathrm{QR}})$ as the weighted sum of a martingale difference sequence. The detailed proof is given in Section 6.

In the recursive update of asymptotic covariance matrix, the initialization of Σ_1 is determined by the choice of the initial estimate in `Online-QR`. For example, if we set the initial estimate as the standard QR estimate based on the first data block $(X_1, \boldsymbol{y}_1)$, then $\Sigma_1 = \tau(1-\tau) f_\epsilon^{-2}(0) \left(X_1^{*\top} X_1^*\right)^{-1}$ [KBJ78].

Theorem 4. *Under the assumptions (i)-(iv) in Theorem 3, the* `Online-QR` *estimator enjoys a sublinear convergence rate*

$$\left\|\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} - \boldsymbol{\beta}^{\mathrm{QR}}\right\|_2 = O_p\left(\frac{1}{t}\right)$$

The estimation error of `Online-QR` can be decomposed as follows:

$$\begin{aligned} \overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} - \boldsymbol{\beta}^{\mathrm{QR}} &= (1-r_t)\left(\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}} - \boldsymbol{\beta}^{\mathrm{QR}}\right) - r_t\left(\boldsymbol{\beta}^{\mathrm{QR}} - \widetilde{\boldsymbol{\beta}}_t^{\mathrm{QR}}\right) \\ &= \left[\prod_{l=1}^{t}(1-r_l)\right]\left(\overline{\boldsymbol{\beta}}_0^{\mathrm{QR}} - \boldsymbol{\beta}^{\mathrm{QR}}\right) - \sum_{l=1}^{t} r_l \left[\prod_{m=l+1}^{t}(1-r_m)\right]\left(\boldsymbol{\beta}^{\mathrm{QR}} - \widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}}\right) \end{aligned} \tag{25}$$

where $r_t := \frac{n_t}{\sum_{l=1}^t n_l}$. In our proof, the initial error due to $\overline{\boldsymbol{\beta}}_0^{\mathrm{QR}}$ is controlled by quantifying the multiplying factor in front of it, while the additional errors from $\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}}$'s occurring in the following updates can be controlled by the Pinelis-Bernstein inequality for martingale difference sequences [TY14]. Then we are able to prove the linear convergence rate, typically true for optimization problems with smooth objectives [MLW+20], without imposing more strict assumptions. The detailed proof is given in Section 6. Besides, it is worth noting that in the proof we do not require $n_t \to \infty$, but only conditions on the homogeneity of data block sizes, i.e., $\max_t n_t / \min_t n_t = O(1)$, which is different from other existing methods based on divide-and-conquer techniques, e.g., [ZDW13].

In the analysis of online learning algorithms, regret is an important statistical risk measure that evaluates the predictive performance of the algorithm. In Theorem 5, we give the regret growth rate of our `Online-QR`.

Theorem 5. *Under the assumptions (i)-(iv) in Theorem 3, the regret*

$$R_t(\boldsymbol{\omega}) = \sum_{l=1}^{t} \frac{1}{n_l}\left[\sum_{i=1}^{n_l} \rho_\tau(y_{il} - {x_{il}^*}^\top \overline{\boldsymbol{\beta}}_l^{\mathrm{QR}}) - \sum_{i=1}^{n_l} \rho_\tau(y_{il} - {x_{il}^*}^\top \boldsymbol{\omega})\right] \qquad \boldsymbol{\omega} \in \mathbb{R}^{p+1}$$

of the `Online-QR` *learning procedure grows at a rate of* $O(\sqrt{t})$*.*

For the online learning procedure with Lipschitz loss, [CBO21] showed that the lower bound of the regret growth rate is $O(\sqrt{t})$. By utilizing the additive inequality of the check loss function (see Lemma 1 in [FLL23]), we can prove the regret of our `Online-QR` learning procedure also grows at the rate of $O(\sqrt{t})$.

5.1 Future directions

Our proposed methods may be further extended to several promising directions, including but not limited to:

(i) We currently estimate coefficients in the normal linear regression model (12) by ordinary least squares (OLS) for its simplicity and computational efficiency. In general, any consistent linear regression estimation techniques can be considered. For example, it is tempting to use weighted least squares (WLS) for the heterogeneity in the model's error term. However, we found using the WLS resulted in lower estimation accuracy compared to the OLS. The choice of the regression estimator still deserves more careful study in our future work. Second, possible improvement might be done to the choice of the penalty parameter λ_t in the nonzero centered ridge regression (19). In this paper, we set λ_t as a function of the data block size to guarantee the problem has a closed-form solution. Potentially, a better calibrated penalty parameter, e.g., letting λ_t reflects the uncertainty in the `Online-QR` estimate at time point $(t-1)$, may further improve the estimation accuracy of `Online-QR`. Third, `Online-QR` might give conservative confidence intervals at the extreme quantile $\tau = 0.9$. In the QR literature, this is a situation that often requires special treatment, such as using extreme value theory in [CFV11] or the correction in [YWH16].

(ii) As the idea in our proposed `Online-QR` is fairly general, it may possibly further extend to other variants of QR, such as censored QR [YNH18, HPTZ22] and Tobit QR [YS07, JLZ12]. Another intriguing topic is to consider `Online-QR` for high-dimensional QR analysis of stream data. However, as sparsity is usually assumed for high-dimensional regression, the prior choice would require more careful consideration. Then it is nontrivial how to adapt the Bayesian formulation into the `Online-QR` framework and meanwhile maintain the computational simplicity.

(iii) Although the popular stochastic gradient descent (SGD) is not applicable to QR problems due to the non-smooth objective, it is tempting to consider its variant based on sub-gradient, due to the fact that it does not require any matrix inversion. However, this algorithm can be numerically unstable and sensitive to the choice of error distributions, likely because the sub-gradient does not necessarily correspond to a real gradient ascent direction [YS18]. Alternatively, one may consider applying SGD to a smoothed approximation to the check loss, e.g., the conquer method [HPTZ23]. Potential challenge with this approach is the selection of smoothing parameter when facing the storage and speed challenges in stream data. The idea in [TBZ22] for choosing local smoothing parameter in distributed QR might provide valuable insight on solving this issue. Further exploration along this direction may also prove useful for solving high-dimensional QR and censored QR in stream data based on the works of [HPTZ22], [TWZ22], [MPTZ24], and [MZ23].

6 Deferred proofs

6.1 Proof of Theorem 1

Let $(Y_i)_{i\geq 1}$ be a sequence of i.i.d. and centered random variables. Denote $\sigma^2 = \mathbb{E}Y_1^2$ and $T_n = \sum_{i=1}^n Y_i$. Cramér [Cra38] has established the following asymptotic expansion on the probabilities of moderate deviations for T_n.

Lemma 2. *Assume that* $\mathbf{E}e^{\lambda|Y_1|} < \infty$ *for a constant* $\lambda > 0$. *Then it holds*

$$\log \frac{\mathbb{P}(T_n > x\sigma\sqrt{n})}{1-\Phi(x)} = O\left(\frac{1+x^3}{\sqrt{n}}\right) \quad as \quad n \to \infty, \tag{26}$$

uniformly for $0 \leq x = o(n^{1/2})$.

The Cramér moderate deviations have attracted a lot of interests. We refer to Petrov [Pet75], Beknazaryan, Sang and Xiao [BSX19], Fan, Hu and Ma [FHM21], Fan, Hu and Xu [FHX24] for more such type results.

We first give a proof for the case $R_n(x,p)$. For all $t \geq 0$, it is easy to see that

$$\begin{aligned}\mathbb{P}\big(R_n(x,p) \geq t\big) &= \mathbb{P}\bigg(\frac{\sqrt{n}f(x_p)(x_{n,p}-x_p)}{\sqrt{p(1-p)}} \geq t\bigg) \\ &= \mathbb{P}\bigg(x_{n,p}-x_p \geq \frac{t\sqrt{p(1-p)}}{\sqrt{n}f(x_p)}\bigg)\end{aligned}$$

Write $x_{n,p,t} = x_p + \frac{t\sqrt{p(1-p)}}{\sqrt{n}f(x_p)}$. Then, by the definition of $x_{n,p,t}$, we get for all $t \geq 0$,

$$\begin{aligned}\mathbb{P}\big(R_n(x,p) \geq t\big) &= \mathbb{P}\big(x_{n,p} \geq x_{n,p,t}\big) = \mathbb{P}\big(p \geq F_n(x_{n,p,t})\big) \\ &= \mathbb{P}\Big(np \geq \sum_{i=1}^n \mathbf{1}(X_i \leq x_{n,p,t})\Big) \\ &= \mathbb{P}\Big(np - nF(x_{n,p,t}) \geq \sum_{i=1}^n \big(\mathbf{1}(X_i \leq x_{n,p,t}) - F(x_{n,p,t})\big)\Big) \\ &= \mathbb{P}\bigg(\sum_{i=1}^n \frac{\mathbf{1}(X_i \leq x_{n,p,t}) - F(t_{n,p}(x))}{\sqrt{nF(x_{n,p,t})(1-F(x_{n,p,t}))}} \leq \frac{\sqrt{n}(p-F(x_{n,p,t}))}{\sqrt{F(x_{n,p,t})(1-F(x_{n,p,t}))}}\bigg)\end{aligned}$$

Recall that $F'(x_p) = f(x_p)$ and that $F''(x) = f'(x)$ is bounded in a neighborhood of $x = x_p$. Thus, it holds uniformly for $0 \leq t = o(\sqrt{n})$,

$$\begin{aligned}p - F(x_{n,p,t}) &= F(x_p) - F(x_{n,p,t}) \\ &= -\frac{t}{\sqrt{n}}\sqrt{p(1-p)} + O\Big(\frac{t^2}{n}\Big)\end{aligned}$$

From the last line, we deduce that uniformly for $0 \leq t = o(\sqrt{n})$,

$$F(x_{n,p,t})\big(1-F(x_{n,p,t})\big) = p(1-p) + \frac{(1-2p)t}{\sqrt{np(1-p)}}p(1-p) + O\Big(\frac{t^2}{n}\Big) \tag{27}$$

Hence, we have uniformly for $0 \leq t = o(\sqrt{n})$,

$$\frac{\sqrt{n}(p-F(x_{n,p,t}))}{\sqrt{F(x_{n,p,t})(1-F(x_{n,p,t}))}} = -t + O\bigg(\frac{t^2}{\sqrt{np(1-p)}}\bigg)$$

Therefore, we deduce that uniformly for $0 \leq t = o(\sqrt{n})$,

$$\mathbb{P}\big(R_n(x,p) \geq t\big) = \mathbb{P}\bigg(\sum_{i=1}^n \frac{\mathbf{1}(X_i \leq x_{n,p,t}) - F(t_{n,p}(x))}{\sqrt{nF(x_{n,p,t})(1-F(x_{n,p,t}))}} \leq -t + O\bigg(\frac{t^2}{\sqrt{np(1-p)}}\bigg)\bigg)$$

Denote $Z_i = \mathbf{1}(X_i \leq x_{n,p,t}) - F(t_{n,p}(x)), 1 \leq i \leq n$. Notice that $(Z_i)_{1\leq i\leq n}$ are i.i.d. and centered random variables, and satisfy that for all $1 \leq i \leq n$,

$$|Z_i| \leq 1 \qquad \text{and} \qquad \mathrm{Var}(Z_i) = F(x_{n,p,t})\big(1-F(x_{n,p,t})\big)$$

The last line and (27) implies that

$$\sum_{i=1}^n \mathrm{Var}(Z_i) = nF(x_{n,p,t})\big(1-F(x_{n,p,t})\big) = np(1-p) + O\big(\sqrt{n}\big) \tag{28}$$

By Lemma 2 and (28), we obtain that it holds

$$\log \frac{\mathbb{P}\big(R_n(x,p) \geq t\big)}{1-\Phi(t)} \ = \ O\bigg(\frac{1+t^3}{\sqrt{n}}\bigg) \ \text{ as } \ n \to \infty \tag{29}$$

uniformly for $0 \leq t = o(n^{1/2})$, which gives the desired inequality for $R_n(x,p)$. For $-R_n(x,p)$, the desired inequality follows by a similar argument. □

6.2 Proof of Theorem 2

Recall that $(Y_i)_{i\geq 1}$ is a sequence of i.i.d. and centered random variables and $T_n = \sum_{i=1}^n Y_i$. Assume that $\mathbf{E}e^{\lambda|Y_1|} < \infty$ for a constant $\lambda > 0$. Denote

$$\Lambda^*(x) = \sup_{\lambda \geq 0}\{\lambda x - \log \mathbf{E}e^{\lambda Y_1}\}$$

the Fenchel-Legendre transform of the cumulant function of Y_1. The function $\Lambda^*(x)$ is known as the good rate function in LDP theory, see Dembo and Zeitouni [DZ98]. Bahadur and Rao [BR60] have established the following sharp large deviations.

Lemma 3. *Assume that $\mathbf{E}e^{\lambda|Y_1|} < \infty$ for a constant $\lambda > 0$. For any $y > 0$, let τ_y and σ_y be the positive solutions of the following equations:*

$$h'(\tau_y) = 0 \qquad \textit{and} \qquad \sigma_y = \sqrt{-h''(\tau_y)}$$

where $h(\tau) = \tau y - \log \mathbf{E}e^{\tau Y_1}$. Then for a given positive constant y, it holds

$$\mathbb{P}\bigg(\frac{T_n}{n} > y\bigg) = \frac{e^{-n\,\Lambda^*(y)}}{\sigma_y\,\tau_y\sqrt{2\pi n}}\Big[1+o(1)\Big], \qquad n \to \infty \tag{30}$$

Such type large deviations have attracted a lot of attentions. We refer to Bercu and Rouault [BR02], Joutard [Jou06], Fan, Grama and Liu [FGL15], Li [Li24] for more such type results.

We are in position to prove Theorem 2. For all $t \geq 0$, it holds

$$\begin{aligned}
\mathbb{P}\big(x_{n,p} - x_p \geq t\big) \ &= \ \mathbb{P}\big(x_{n,p} \geq x_p + t\big) \ = \ \mathbb{P}\big(p \geq F_n(x_p+t)\big) \\
&= \ \mathbb{P}\Big(n(1-p) \leq \sum_{i=1}^n \mathbf{1}(X_i > x_p + t)\Big) \\
&= \ \mathbb{P}\bigg(n(F(x_p+t) - p) \leq \sum_{i=1}^n U_i\bigg)
\end{aligned}$$

where

$$U_i = \mathbf{1}(X_i > x_p + t) - 1 + F(x_p+t), \qquad i = 1, ..., n$$

Notice that $(U_i)_{1\leq i\leq n}$ are i.i.d. and centered random variables with $|U_i| \leq 1$. By some simple calculations, it is easy to see that

$$\begin{aligned}
\mathbf{E}e^{\lambda U_1} \ &= \ e^{\lambda F(x_p+t)}\big(1 - F(x_p+t)\big) + e^{\lambda(F(x_p+t)-1)}F(x_p+t) \\
&= \ \exp\Big\{\lambda F(x_p+t) + \log\Big(1 - F(x_p+t) + e^{-\lambda}F(x_p+t)\Big)\Big\}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}\Lambda^*(F(x_p+t)-p) &= \sup_{\lambda\ge 0}\Big\{\lambda(F(x_p+t)-p)-\log \mathbf{E}e^{\lambda U_1}\Big\}\\ &= \sup_{\lambda\ge 0}\Big\{-\lambda p-\log\Big(1-F(x_p+t)+e^{-\lambda}F(x_p+t)\Big)\Big\}\\ &= \Lambda^+(t)\end{aligned}$$

where

$$\Lambda^+(t)=p\log\frac{p}{F(x_p+t)}+(1-p)\log\frac{1-p}{1-F(x_p+t)}$$

Denote

$$h_1(\tau)=\tau(F(x_p+t)-p)-\log \mathbf{E}e^{\tau U_1}$$

Let τ_t^+ and σ_t^+ be the positive solutions of the following equations:

$$h_1'(\tau_t^+)=0 \quad\text{ and }\quad \sigma_t^+=\sqrt{-h_1''(\tau_t^+)}$$

Then we have

$$\tau_t^+=\log\frac{F(x_p+t)(1-p)}{p(1-F(x_p+t))}\quad\text{ and }\quad \sigma_t^+=\sigma_p=\sqrt{p(1-p)}$$

Applying Lemma 3 to $(U_i)_{1\le i\le n}$, with

$$Y_i=U_i,\quad y=F(x_p+t)-p,\quad \tau_t=\tau_t^+\quad\text{ and }\quad \sigma_t=\sigma_p$$

we get for any $t>0$,

$$\mathbb{P}\big(x_{n,p}-x_p\ge t\big)=\frac{1}{\sigma_p\,\tau_t^+\sqrt{2\pi n}}\exp\Big\{-n\Lambda^+(t)\Big\}\Big[1+o(1)\Big],\quad n\to\infty$$

which gives the first desired equality. An argument in symmetry gives the second desired equality.

6.3 Proof of Theorem 3

We first show that the `Online-QR` estimator can achieve the unbiasedness after finite time, and then prove its asymptotic normality as $t\to\infty$. Finally, we give a recursive formula for the update of its asymptotic covariance matrix.

6.3.1 The unbiasedness of the `Online-QR` estimator

From the update rule of `Online-QR`, we see that $\Big\{\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}\mid t=1,2,\dots\Big\}$ actually forms a time-homogeneous Markov process. According to Theorem 8.2.14 in [Sta09], under the following three conditions: (i) the irreducibility of the process; (ii) the geometric drift of the process to the center; (iii) the uniform integrability of the marginal densities of the process, a time-homogeneous Markov process is stationary. Condition (iii) follows from the general argument laid out in [Sta09]. Next, we first verify conditions (i) and (ii), and then prove the unbiasedness of `Online-QR` under the stationarity of this Markov process.

From (19), $\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}$ is obtained by solving a penalized linear regression problem, and thus the state space S of the process is a compact subset of $\mathbb{R}^{p+1}$. To verify the irreducibility, we only need to show that with positive probability any $\boldsymbol{\beta}' \in S$ is reachable from any $\boldsymbol{\beta}'' \in S$ after finite time. Note that $\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}$ can be rewritten as

$$\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} = \frac{n_t}{\sum_{l=1}^t n_l}(X_t^{*\top} X_t^*)^{-1}\left[X_t^{*\top}\boldsymbol{y}_t^* + \frac{\sum_{l=1}^{t-1} n_l}{n_t} X_t^{*\top} X_t^* \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right]$$

from which we see that its value is constrained to the subspace spanned by the rows of X_t^*. Therefore, if $\bigcap_{l=t_0}^{t_0+t_0'} \mathrm{null}(X_l^*) = \mathbf{0}_{p+1}$ for $t_0, t' \in \mathbb{N}$ and t' is large enough, any value in S is reachable after finite time. Here $\mathrm{null}(X_l^*)$ represents the null space of the linear map induced by X_l^*. While if $\bigcap_{l=t_0}^{t_0+t'} \mathrm{null}(X_l^*) \neq \mathbf{0}_{p+1}$, the state space of the process is reducible to $S' = S \backslash \bigcap_{l=t_0}^{t_0+t'} \mathrm{null}(X_l^*)$, and any value in S' is reachable after finite time. To assess the geometric drift of the process to the center, we utilize

$$\begin{aligned}\left\|\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}\right\|_2 &= \left\|\frac{n_t}{\sum_{l=1}^t n_l}\widetilde{\boldsymbol{\beta}}_t^{\mathrm{QR}} + \frac{\sum_{l=1}^{t-1} n_l}{\sum_{l=1}^t n_l}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right\|_2 \\ &\leq \frac{n_t}{\sum_{l=1}^t n_l}\left\|\widetilde{\boldsymbol{\beta}}_t^{\mathrm{QR}}\right\|_2 + \frac{\sum_{l=1}^{t-1} n_l}{\sum_{l=1}^t n_l}\left\|\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right\|_2 \leq \frac{c_1}{t} + c_2\left\|\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right\|_2\end{aligned} \tag{31}$$

where $c_1, c_2 \in \mathbb{R}$ are positive constants and $c_2 < 1$. From (31), we can conclude the tightness of $\{\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} : t = 1, 2, \ldots\}$. This then implies that, after finite time t_0, the process can reach the stationarity.

Once the stationarity is reached, we have $\mathbb{E}(\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}) = \mathbb{E}(\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}})$. In what follows, we denote this expectation as $\boldsymbol{\beta}^*$ and will show that it is equal to the true value of the regression parameter $\boldsymbol{\beta}$. By (17) and (20), we have

$$\begin{aligned}\boldsymbol{\beta}^* &= \mathbb{E}\left[\frac{n_t}{\sum_{l=1}^t n_l}(X_t^{*\top} X_t^*)^{-1} X_t^{*\top}\boldsymbol{y}_t^* + \frac{\sum_{l=1}^{t-1} n_l}{\sum_{l=1}^t n_l}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right] \\ &= \frac{n_t}{\sum_{l=1}^t n_l}\mathbb{E}\left[(X_t^{*\top} X_t^*)^{-1} X_t^{*\top}\boldsymbol{y}_t^*\right] + \frac{\sum_{l=1}^{t-1} n_l}{\sum_{l=1}^t n_l}\boldsymbol{\beta}^* \\ &= \frac{\xi_1 n_t}{|\widehat{\gamma}|\xi_2\sum_{l=1}^t n_l}(X_t^{*\top} X_t^*)^{-1} X_t^{*\top}\left(\frac{1-2\tau+2\tau^2}{\tau(1-\tau)}\mathbf{1}_{n_t} - \mathbb{E}|\boldsymbol{y}_t - X_t^*\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right) + \frac{n_t}{\sum_{l=1}^t n_l}\boldsymbol{\beta} + \frac{\sum_{l=1}^{t-1} n_l}{\sum_{l=1}^t n_l}\boldsymbol{\beta}^*\end{aligned} \tag{32}$$

where $\mathbf{1}_{n_t}$ and $\left|y_t - X_t^*\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right|$ respectively are the n_t-dimensional vectors composed by 1 and $|y_{it} - x_{it}^{*\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|$. For the i th element of $\mathbb{E}\left|y_t - X_t^*\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right|$, we have

$$\begin{aligned}\mathbb{E}|y_{it} - x_{it}^{*\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}| &= \mathbb{E}\left[\mathbb{E}\left(|y_{it} - x_{it}^{*\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}| \mid y_{it}\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(\left|y_{it} - x_{it}^{*\top}\boldsymbol{\beta}^*\right| + \nabla h\left(\boldsymbol{\beta}'\right)^\top\left(\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}} - \boldsymbol{\beta}^*\right) \mid y_{it}\right)\right] = \mathbb{E}\left|y_{it} - \boldsymbol{x}_{it}^{*\top}\boldsymbol{\beta}^*\right|\end{aligned}$$

where $\nabla h\left(\boldsymbol{\beta}'\right)$ is the sub-gradient of the absolute value function at $\boldsymbol{\beta}'$, and $\boldsymbol{\beta}'$ lies between $\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}$ and $\boldsymbol{\beta}^*$. Under the assumption $\epsilon_{it} \sim \mathrm{AL}(0, 1, \tau)$, the variable $w_{it} = y_{it} - x_{it}^{*\top}\boldsymbol{\beta}^* = x_{it}^{*\top}(\boldsymbol{\beta} - \boldsymbol{\beta}^*) + \epsilon_{it}$ follows $\mathrm{AL}\left(x_{it}^{*\top}(\boldsymbol{\beta} - \boldsymbol{\beta}^*), 1, \tau\right)$ with density $f(w_{it}) = \tau(1-\tau)\exp\left\{-\rho_\tau(w_{it} - \boldsymbol{x}_{it}^{*\top}(\boldsymbol{\beta} - \boldsymbol{\beta}^*))\right\}$. Then we have

$$\mathbb{E}|y_{it} - x_{it}^{*\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}| = \mathbb{E}|w_{it}| = \int_0^{+\infty} w_{it} f(w_{it}) d_{w_{it}} - \int_{-\infty}^0 w_{it} f(w_{it}) d_{w_{it}} \tag{33}$$

Next, we evaluate the integrals in the right-hand side by separating into two cases, (i) $x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \geq 0$ and (ii) $x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \leq 0$. As the evaluations are done in a similar way, we mainly focus on case (i) in the following. The first integral in (33) can be written as

$$\int_0^{+\infty} w_{it} f(w_{it}) d_{w_{it}} = \int_0^{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)} w_{it} f(w_{it}) d_{w_{it}} + \int_{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)}^{+\infty} w_{it} f(w_{it}) d_{w_{it}}$$

where the first term in the right-hand side is

$$\begin{aligned}
&\int_0^{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)} w_{it} f(w_{it}) d_{w_{it}} \\
&= \int_0^{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)} w_{it}\tau(1-\tau)\exp\left\{(1-\tau)\left[w_{it} - x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\} d_{w_{it}} \\
&= \tau\left\{ w_{it}\exp\left\{(1-\tau)\left[w_{it} - x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\}\Big|_0^{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)} - \int_0^{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)} \exp\left\{(1-\tau)\left[w_{it} - \boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\} d_{w_{it}} \right\} \\
&= \tau\left\{ x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) - \frac{1}{1-\tau} + \frac{1}{1-\tau}\exp\left\{(\tau-1)x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right\}\right\}
\end{aligned}$$

and the second term is

$$\begin{aligned}
&\int_{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)}^{+\infty} w_{it} f(w_{it}) d_{w_{it}} = \int_{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)}^{+\infty} w_{it}\tau(1-\tau)\exp\left\{-\tau\left[w_{it} - x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\} d_{w_{it}} \\
&= (\tau-1)\left\{ w_{it}\exp\left\{-\tau\left[w_{it} - x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\}\Big|_{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)}^{+\infty} \right. \\
&\qquad\qquad \left. - \int_{x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)}^{+\infty} \exp\left\{-\tau\left[w_{it} - x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\} d_{w_{it}} \right\} \\
&= (1-\tau)\left\{ x_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) + \frac{1}{\tau}\right\}
\end{aligned}$$

Thus

$$\int_0^{+\infty} w_{it} f(w_{it}) d_{w_{it}} = \boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) + \xi_1 + \frac{\tau}{1-\tau}\exp\left\{(\tau-1)\boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right\}$$

Meanwhile, the second integral in (33) is

$$\begin{aligned}
&\int_{-\infty}^{0} w_{it} f(w_{it}) d_{w_{it}} = \int_{-\infty}^{0} w_{it}\tau(1-\tau)\exp\left\{(1-\tau)\left[w_{it} - \boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\} d_{w_{it}} \\
&= \tau\left\{ w_{it}\exp\left\{(1-\tau)\left[w_{it} - \boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\}\Big|_{-\infty}^{0} - \int_{-\infty}^{0} \exp\left\{(1-\tau)\left[w_{it} - \boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right]\right\} d_{w_{it}} \right\} \\
&= -\frac{\tau}{1-\tau}\exp\left\{(\tau-1)\boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right\}
\end{aligned}$$

Therefore, we have

$$\mathbb{E}\left| y_{it} - \boldsymbol{x}_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right| = \boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) + \xi_1 + \frac{2\tau}{1-\tau}\exp\left\{(\tau-1)\boldsymbol{x}_{it}^{*}{}^{\top}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right\} \tag{34}$$

On the other hand, if ${x_{it}^*}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \leq 0$, it is easy to find using similar techniques that

$$\mathbb{E}\left|y_{it}-{\boldsymbol{x}_{it}^*}^\top \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right| = -{x_{it}^*}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^*) - \xi_1 + \frac{2(1-\tau)}{\tau}\exp\left\{\tau {x_{it}^*}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right\} \tag{35}$$

Substituting (34) and (35) into (32), and utilizing the fact that $0<\tau<1$, $\exp\left\{(\tau-1){\boldsymbol{x}_{it}^*}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right\} \leq 1$ for ${\boldsymbol{x}_{it}^*}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \geq 0$, and $\exp\left\{\tau{\boldsymbol{x}_{it}^*}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right\} \leq 1$ for ${\boldsymbol{x}_{it}^*}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \leq 0$, we then finally obtain $\boldsymbol{\beta}^* = \boldsymbol{\beta}$.

6.3.2 The asymptotic normality of the `Online-QR` estimator

By the update rule of `Online-QR`, we can write $\left(\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} - \boldsymbol{\beta}\right)$ as

$$\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} - \boldsymbol{\beta} = \sum_{l=2}^{t}\frac{n_l}{\sum_{m=1}^t n_m}(\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}} - \boldsymbol{\beta}) + \frac{n_1}{\sum_{m=1}^t n_m}(\overline{\boldsymbol{\beta}}_1^{\mathrm{QR}} - \boldsymbol{\beta})$$

By (34) and (35), we have

$$\begin{aligned}
&\mathbb{E}\left(\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}} - \boldsymbol{\beta} \mid \overline{\boldsymbol{\beta}}_1^{\mathrm{QR}}, \widetilde{\boldsymbol{\beta}}_2^{\mathrm{QR}}, \ldots, \widetilde{\boldsymbol{\beta}}_{(l-1)}^{\mathrm{QR}}\right)\\
&= \mathbb{E}\left[({X_l^*}^\top X_l^*)^{-1}{X_l^*}^\top\left(X_l^*\boldsymbol{\beta} + \epsilon_l - \xi_1\widetilde{\boldsymbol{v}}_l\right) - \boldsymbol{\beta} \mid \overline{\boldsymbol{\beta}}_1^{\mathrm{QR}}, \widetilde{\boldsymbol{\beta}}_2^{\mathrm{QR}}, \ldots, \widetilde{\boldsymbol{\beta}}_{(l-1)}^{\mathrm{QR}}\right]\\
&= ({X_l^*}^\top X_l^*)^{-1}{X_l^*}^\top\left[\frac{1-2\tau+2\tau^2}{\tau(1-\tau)}\mathbf{1}_{n_t} - \mathbb{E}\left(\left|\boldsymbol{y}_l - X_l^*\overline{\boldsymbol{\beta}}_{(l-1)}^{\mathrm{QR}}\right| \overline{\boldsymbol{\beta}}_1^{\mathrm{QR}}, \widetilde{\boldsymbol{\beta}}_2^{\mathrm{QR}}, \ldots, \widetilde{\boldsymbol{\beta}}_{(l-1)}^{\mathrm{QR}}\right)\right]\\
&= \mathbf{0}_{p+1} \qquad l = 2,3,\ldots
\end{aligned}$$

where $\epsilon_l = (\epsilon_{1l}, \epsilon_{2l}, \ldots, \epsilon_{n_l l})^\top \in \mathbb{R}^{n_l}$. Then if the initialization of `Online-QR` is chosen as the standard QR estimator based on $(X_1, \boldsymbol{y}_1)$, $(\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}} - \boldsymbol{\beta})(l = 1,2,\ldots)$ is a martingale difference sequence. Here we write $\overline{\boldsymbol{\beta}}_1^{\mathrm{QR}}$ as $\widetilde{\boldsymbol{\beta}}_1^{\mathrm{QR}}$ for simplicity. Thus, we can use the martingale central limit theorem to prove the asymptotic normality. Next, we verify the Lindeberg-type condition.

For any $z \in \mathbb{R}^{p+1}$ satisfying $\|z\|_2 = 1$ and $\alpha > 0$, we have

$$\begin{aligned}
&\mathbb{E}\left|\boldsymbol{z}^\top(\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}} - \boldsymbol{\beta})\right|^{2+\alpha}\\
&= \mathbb{E}\left|\boldsymbol{z}^\top\left[({X_l^*}^\top X_l^*)^{-1}{X_l^*}^\top\left(\xi_1\boldsymbol{v}_l + \xi_2\sqrt{\boldsymbol{v}_l}\circ\boldsymbol{u}_l - \xi_1\widetilde{\boldsymbol{v}}_l\right)\right]\right|^{2+\alpha}\\
&\leq \left(2\lambda_{\min}^{-1}\sqrt{\lambda_{\max}}/\sqrt{n_l}\right)^{2+\alpha}\left[\mathbb{E}\left\|\xi_1\left(\boldsymbol{v}_l - \widetilde{\boldsymbol{v}}_l\right)\right\|_2^{2+\alpha} + \mathbb{E}\left\|\xi_2\sqrt{\boldsymbol{v}_l}\circ\boldsymbol{u}_l\right\|_2^{2+\alpha}\right]\\
&\leq c_3\left(2\lambda_{\min}^{-1}\sqrt{\lambda_{\max}}\right)^{2+\alpha}
\end{aligned} \tag{36}$$

where $c_3 \in \mathbb{R}$ is a constant, "$\circ$" represents the Hadamard product, and $u_l = (u_{1l}, u_{2l}, \ldots, u_{n_l l}) \in \mathbb{R}^{n_l}$. In (36), we utilize

$$\mathbb{E}\left(v_{il} - \widetilde{v}_{il}\right) = \frac{1-2\tau+2\tau^2}{\tau(1-\tau)} - \mathbb{E}\left|y_{il} - {x_{il}^*}^\top\overline{\boldsymbol{\beta}}_{(l-1)}^{\mathrm{QR}}\right| = 0 \tag{37a}$$

$$\mathbb{E}\left(\sqrt{v_{il}}u_{il}\right) = \mathbb{E}\left(\sqrt{v_{il}}\right)\mathbb{E}\left(u_{il}\right) = 0 \tag{37b}$$

$$\mathrm{Var}\left(v_{il} - \widetilde{v}_{il}\right) \leq 2\,\mathrm{Var}\left(v_{il}\right) + 2\,\mathrm{Var}\left(\widetilde{v}_{il}\right) = 2 + 2\frac{\xi_1^2}{\widehat{\gamma}^2\xi_2^2}\mathrm{Var}\left(\left|y_{il} - {x_{il}^*}^\top\overline{\boldsymbol{\beta}}_{(l-1)}^{\mathrm{QR}}\right|\right) = O(1) \tag{37c}$$

and

$$\operatorname{Var}\left(\sqrt{v_{il}} u_{il}\right)=\mathbb{E}\left(\operatorname{Var}\left(\sqrt{v_{il}} u_{il} \mid v_{il}\right)\right)+\operatorname{Var}\left(\mathbb{E}\left(\sqrt{v_{il}} u_{il} \mid v_{il}\right)\right)=1 \tag{38}$$

The specific form of $\operatorname{Var}\left(\left|y_{il}-x_{il}^{*}{}^{\top} \overline{\boldsymbol{\beta}}_{(l-1)}^{\mathrm{QR}}\right|\right)$ is given in (42) in the third step of this proof. By (36), we can further obtain

$$\begin{aligned}
& \frac{1}{t} \sum_{l=1}^{t} \mathbb{E}\left[\left|z^{\top}(\widetilde{\boldsymbol{\beta}}_{l}^{\mathrm{QR}}-\boldsymbol{\beta})\right|^{2} \mathbb{I}\left(\left|\boldsymbol{z}^{\top}(\widetilde{\boldsymbol{\beta}}_{l}^{\mathrm{QR}}-\boldsymbol{\beta})\right|>\varepsilon \sqrt{t}\right)\right] \\
& \leq \frac{1}{t} \sum_{l=1}^{t} \varepsilon^{-\alpha} t^{-\frac{\alpha}{2}} \mathbb{E}\left[\left|\boldsymbol{z}^{\top}(\widetilde{\boldsymbol{\beta}}_{l}^{\mathrm{QR}}-\boldsymbol{\beta})\right|^{2+\alpha} \mathbb{I}\left(\left|z^{\top}(\widetilde{\boldsymbol{\beta}}_{l}^{\mathrm{QR}}-\boldsymbol{\beta})\right|>\varepsilon \sqrt{t}\right)\right] \\
& \leq \varepsilon^{-\alpha} t^{-\frac{\alpha}{2}-1} \sum_{l=1}^{t} \mathbb{E}\left|\boldsymbol{z}^{\top}(\widetilde{\boldsymbol{\beta}}_{l}^{\mathrm{QR}}-\boldsymbol{\beta})\right|^{2+\alpha} \rightarrow 0 \quad \text{ as } \quad t \rightarrow \infty
\end{aligned}$$

for any $\varepsilon>0$, and thus $\overline{\boldsymbol{\beta}}_{t}^{\mathrm{QR}}$ is asymptotically normal as we have $\frac{n_{l}}{\sum_{m=1}^{t} n_{m}}=O\left(\frac{1}{t}\right)$ under the assumption $\max_{l} n_{l} / \min_{l} n_{l}=O(1)$.

6.3.3 The asymptotic covariance matrix of the `Online-QR` estimator

Denote $\operatorname{Var}(\overline{\boldsymbol{\beta}}_{t}^{\mathrm{QR}})$ as $\boldsymbol{\Sigma}_{t}$. By (17) and (20), we have

$$\boldsymbol{\Sigma}_{t}=\operatorname{Var}\left(\mathbb{I}_{1}\right)+\operatorname{Var}\left(\mathbb{I}_{2}\right)+\operatorname{Var}\left(\mathbb{I}_{3}\right)-2 \operatorname{Cov}\left(\mathbb{I}_{1}+\mathbb{I}_{2}, \mathbb{I}_{3}\right) \tag{39}$$

where $\mathbb{I}_{1}=\frac{n_{t}}{\sum_{l=1}^{t} n_{l}}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1} X_{t}^{*}{}^{\top} \boldsymbol{y}_{t}, \mathbb{I}_{2}=\frac{\sum_{l=1}^{t-1} n_{l}}{\sum_{l=1}^{t} n_{l}} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}$ and $\mathbb{I}_{3}=\frac{n_{t} \xi_{1}}{\sum_{l=1}^{t} n_{l}}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1} X_{t}^{*}{}^{\top} \widetilde{\boldsymbol{v}}_{t}$. For the first two terms on the right-hand side of (39), we have

$$\begin{aligned}
\operatorname{Var}\left(\mathbb{I}_{1}\right) & =\frac{n_{t}^{2}}{\left(\sum_{l=1}^{t} n_{l}\right)^{2}} \cdot \frac{1-2 \tau+2 \tau^{2}}{\tau^{2}(1-\tau)^{2}}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1} X_{t}^{*}{}^{\top} X_{t}^{*}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1} \\
& =\frac{n_{t}^{2}}{\left(\sum_{l=1}^{t} n_{l}\right)^{2}} \cdot \frac{1-2 \tau+2 \tau^{2}}{\tau^{2}(1-\tau)^{2}}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1}
\end{aligned} \tag{40}$$

and

$$\operatorname{Var}\left(\mathbb{I}_{2}\right)=\frac{\left(\sum_{l=1}^{t-1} n_{l}\right)^{2}}{\left(\sum_{l=1}^{t} n_{l}\right)^{2}} \Sigma_{t-1} \tag{41}$$

while the third term on the right-hand side of (39) is

$$\begin{aligned}
\operatorname{Var}\left(\mathbb{I}_{3}\right) & =\frac{n_{t}^{2} \xi_{1}^{2}}{\left(\sum_{l=1}^{t} n_{l}\right)^{2}}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1} X_{t}^{*}{}^{\top} \operatorname{Var}\left(\widetilde{\boldsymbol{v}}_{t}\right) X_{t}^{*}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1} \\
& =\frac{n_{t}^{2}}{\left(\sum_{l=1}^{t} n_{l}\right)^{2}} \cdot \frac{\xi_{1}^{2}}{\widehat{\gamma}^{2} \xi_{2}^{2}}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1} X_{t}^{*}{}^{\top} \operatorname{Var}(|\boldsymbol{y}_{t}-X_{t}^{*} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|) X_{t}^{*}(X_{t}^{*}{}^{\top} X_{t}^{*})^{-1}
\end{aligned}$$

Next, we first analyze the diagonal elements of $\operatorname{Var}(|\boldsymbol{y}_{t}-X_{t}^{*} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|)$, and then give its off-diagonal elements. For the i th diagonal element, we have

$$\operatorname{Var}\left(\left|y_{it}-x_{it}^{*}{}^{\top} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right|\right)=\mathbb{E}\left(y_{it}-x_{it}^{*}{}^{\top} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right)^{2}-\mathbb{E}^{2}\left(\left|y_{it}-x_{t}^{*}{}^{\top} \overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right|\right)$$

in which

$$\begin{aligned}
&\mathbb{E}(y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}})^2 = \mathbb{E}(y_{it})^2 + \mathbb{E}(x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}})^2 - 2\mathbb{E}(y_{it})\mathbb{E}(x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}) \\
&= \mathrm{Var}(y_{it}) + \mathbb{E}^2(y_{it}) + \mathrm{Var}(x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}) + \mathbb{E}^2(x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}) - 2\left(x_{it}^{*}{}^{\top}\boldsymbol{\beta} + \xi_1\right)x_{it}^{*}{}^{\top}\boldsymbol{\beta} \\
&= \frac{1-2\tau+2\tau^2}{\tau^2(1-\tau)^2} + \left(x_{it}^{*}{}^{\top}\boldsymbol{\beta} + \xi_1\right)^2 + x_{it}^{*}{}^{\top}\Sigma_{t-1}x_{it}^{*} + \left(x_{it}^{*}{}^{\top}\boldsymbol{\beta}\right)^2 - 2\left(x_{it}^{*}{}^{\top}\boldsymbol{\beta} + \xi_1\right)x_{it}^{*}{}^{\top}\boldsymbol{\beta} \\
&= \frac{2-6\tau+6\tau^2}{\tau^2(1-\tau)^2} + x_{it}^{*}{}^{\top}\boldsymbol{\Sigma}_{t-1}x_{it}^{*}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left(\left|y_{it} - \boldsymbol{x}_{t}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right|\right) = \mathbb{E}\left[\mathbb{E}\left(\left|y_{it} - x_{t}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right| y_{it}\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\left|y_{it} - x_{t}^{*}{}^{\top}\boldsymbol{\beta}\right| + \nabla h\left(\boldsymbol{\beta}'\right)^{\top}\left(\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}} - \boldsymbol{\beta}\right) \mid y_{it}\right)\right] = \mathbb{E}\left|\epsilon_{it}\right| \\
&= \int_0^{+\infty} \epsilon_{it} f\left(\epsilon_{it}\right) d_{\epsilon_{it}} - \int_{-\infty}^{0} \epsilon_{it} f\left(\epsilon_{it}\right) d_{\epsilon_{it}} \\
&= \int_0^{+\infty} \epsilon_{it}\tau(1-\tau)\exp\left\{-\tau\epsilon_{it}\right\} d_{\epsilon_{it}} - \int_{-\infty}^{0} \epsilon_{it}\tau(1-\tau)\exp\left\{(1-\tau)\epsilon_{it}\right\} d_{\epsilon_{it}} \\
&= (\tau-1)\left[\epsilon_{it}\exp\left\{-\tau\epsilon_{it}\right\}\Big|_0^{+\infty} - \int_0^{+\infty}\exp\left\{-\tau\epsilon_{it}\right\} d_{\epsilon_{it}}\right] \\
&\quad - \tau\left[\epsilon_{it}\exp\left\{(1-\tau)\epsilon_{it}\right\}\Big|_{-\infty}^{0} - \int_{-\infty}^{0}\exp\left\{(1-\tau)\epsilon_{it}\right\} d_{\epsilon_{it}}\right] = \frac{1-2\tau+2\tau^2}{\tau(1-\tau)}
\end{aligned}$$

Thus

$$\mathrm{Var}\left(|y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right) = x_{it}^{*}{}^{\top}\boldsymbol{\Sigma}_{t-1}x_{it}^{*} + \frac{1-2\tau-2\tau^2+8\tau^3-4\tau^4}{\tau^2(1-\tau)^2} \tag{42}$$

For the (i,j) th off-diagonal element of $\mathrm{Var}(|\boldsymbol{y}_t - X_t^{*}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|)$, we have

$$\begin{aligned}
&\mathrm{Cov}\left(|y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|, |y_{jt} - x_{jt}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right) \\
&= \mathbb{E}\left[\mathrm{Cov}\left(\sqrt{\left(y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right)^2}, \sqrt{\left(y_{jt} - x_{jt}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right)^2} \mid y_{it}, y_{jt}\right)\right] \\
&\quad + \mathbb{E}\left\{\left[\mathbb{E}\left(|y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}| \mid y_{it}\right) - \mathbb{E}|y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right]\right. \\
&\qquad \left.\cdot\left[\mathbb{E}\left(|y_{jt} - x_{jt}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}| \mid y_{jt}\right) - \mathbb{E}|y_{jt} - x_{jt}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right]\right\}
\end{aligned}$$

in which

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{Cov}\left(\sqrt{\left(y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right)^2}, \sqrt{\left(y_{jt} - x_{jt}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right)^2} \mid y_{it}, y_{jt}\right)\right] \\
&= \mathbb{E}\left[\mathrm{Cov}\left(x_{it}^{*}{}^{\top}\left(\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}} - \boldsymbol{\beta}\right), \boldsymbol{x}_{jt}^{*}{}^{\top}\left(\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}} - \boldsymbol{\beta}\right) \mid y_{it}, y_{jt}\right)\right] = \boldsymbol{x}_{it}^{*}{}^{\top}\boldsymbol{\Sigma}_{t-1}x_{jt}^{*}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left\{\left[\mathbb{E}\left(|y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}| \mid y_{it}\right) - \mathbb{E}|y_{it} - x_{it}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right]\right. \\
&\qquad \left.\left[\mathbb{E}\left(|y_{jt} - x_{jt}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}| \mid y_{jt}\right) - \mathbb{E}|y_{jt} - x_{jt}^{*}{}^{\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right]\right\} \\
&= \mathbb{E}\left[\left(|\varepsilon_{it}| - \mathbb{E}\left|\epsilon_{it}\right|\right)\left(|\epsilon_{jt}| - \mathbb{E}\left|\epsilon_{jt}\right|\right)\right] = 0
\end{aligned}$$

Thus

$$\operatorname{Cov}\left(|y_{it}-x_{it}^{*\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|,|y_{jt}-x_{jt}^{*\top}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right)=x_{it}^{*\top}\Sigma_{t-1}x_{jt}^{*} \tag{43}$$

Combining (42) and (43), we obtain

$$\operatorname{Var}(|\boldsymbol{y}_t-X_t^*\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|)=X_t^*\boldsymbol{\Sigma}_{t-1}X_t^{*\top}+\frac{1-2\tau-2\tau^2+8\tau^3-4\tau^4}{\tau^2(1-\tau)^2}I_{n_t}$$

and further

$$\operatorname{Var}(\mathbb{I}_3)=\frac{n_t^2(1-2\tau)^2}{\left(\sum_{l=1}^t n_l\right)^2}\left[\Sigma_{t-1}+\frac{1-2\tau-2\tau^2+8\tau^3-4\tau^4}{\tau^2(1-\tau)^2}(X_t^{*\top}X_t^*)^{-1}\right] \tag{44}$$

Finally, for the fourth term in the right-hand side of (39), we have

$$\operatorname{Cov}(\mathbb{I}_1+\mathbb{I}_2,\mathbb{I}_3)=\mathbb{E}\left[\operatorname{Cov}(\mathbb{I}_1+\mathbb{I}_2,\mathbb{I}_3\mid\boldsymbol{y}_t)\right]-\mathbb{E}\left\{\left[\mathbb{E}(\mathbb{I}_1+\mathbb{I}_2\mid\boldsymbol{y}_t)-\mathbb{E}(\mathbb{I}_1+\mathbb{I}_2)\right]\left[\mathbb{E}(\mathbb{I}_3\mid\boldsymbol{y}_t)-\mathbb{E}(\mathbb{I}_3)\right]^\top\right\}$$

in which

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{Cov}(\mathbb{I}_1+\mathbb{I}_2,\mathbb{I}_3\mid\boldsymbol{y}_t)\right]\\
=&\mathbb{E}\left[\operatorname{Cov}\left(\frac{\sum_{l=1}^{t-1}n_l}{\sum_{l=1}^{t}n_l}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}},\frac{n_t\xi_1}{\left(\sum_{l=1}^t n_l\right)|\widehat{\gamma}|\xi_2}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}|\boldsymbol{y}_t-X_t^*\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\boldsymbol{y}_t\right)\right]\\
=&\frac{\left(\sum_{l=1}^{t-1}n_l\right)n_t}{\left(\sum_{l=1}^t n_l\right)^2}\cdot\frac{\xi_1}{|\widehat{\gamma}|\xi_2}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}\mathbb{E}\left[\operatorname{Cov}\left(-\frac{y_t-X_t^*\boldsymbol{\beta}}{|\boldsymbol{y}_t-X_t^*\boldsymbol{\beta}|}\circ X_t^*\left(\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}-\boldsymbol{\beta}\right),\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\mid\boldsymbol{y}_t\right)\right]\\
=&-\frac{\left(\sum_{l=1}^{t-1}n_l\right)n_t}{\left(\sum_{l=1}^t n_l\right)^2}\cdot\frac{\xi_1}{|\widehat{\gamma}|\xi_2}\mathbb{E}\left(\frac{\epsilon}{|\epsilon|}\right)\Sigma_{t-1}=-\frac{\left(\sum_{l=1}^{t-1}n_l\right)n_t(1-2\tau)^2}{\left(\sum_{l=1}^t n_l\right)^2}\Sigma_{t-1}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}(\mathbb{I}_1+\mathbb{I}_2\mid\boldsymbol{y}_t)-\mathbb{E}(\mathbb{I}_1+\mathbb{I}_2)\\
=&\mathbb{E}\left(\frac{n_t}{\sum_{l=1}^t n_l}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}\boldsymbol{y}_t+\frac{\sum_{l=1}^{t-1}n_l}{\sum_{l=1}^t n_l}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\,\middle|\,y_t\right)\\
&-\mathbb{E}\left(\frac{n_t}{\sum_{l=1}^t n_l}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}\boldsymbol{y}_t+\frac{\sum_{l=1}^{t-1}n_l}{\sum_{l=1}^t n_l}\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}\right)\\
=&\frac{n_t}{\sum_{l=1}^t n_l}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}\left[\boldsymbol{y}_t-\mathbb{E}(\boldsymbol{y}_t)\right]=\frac{n_t}{\sum_{l=1}^t n_l}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}(\boldsymbol{\epsilon}_t-\xi_1\mathbf{1}_{n_t})
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}(\mathbb{I}_3\mid\boldsymbol{y}_t)-\mathbb{E}(\mathbb{I}_3)\\
=&\mathbb{E}\left(\frac{n_t\xi_1}{\sum_{l=1}^t n_l}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}\widetilde{\boldsymbol{v}}_t\,\middle|\,\boldsymbol{y}_t\right)-\mathbb{E}\left(\frac{n_t\xi_1}{\sum_{l=1}^t n_l}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}\widetilde{\boldsymbol{v}}_t\right)\\
=&\frac{n_t\xi_1}{\left(\sum_{l=1}^t n_l\right)|\widehat{\gamma}|\xi_2}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}\left[\mathbb{E}\left(|\boldsymbol{y}_t-X_t^*\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\mid\boldsymbol{y}_t\right)-\mathbb{E}|\boldsymbol{y}_t-X_t^*\overline{\boldsymbol{\beta}}_{t-1}^{\mathrm{QR}}|\right]\\
=&\frac{n_t(1-2\tau)}{\left(\sum_{l=1}^t n_l\right)}(X_t^{*\top}X_t^*)^{-1}X_t^{*\top}\left(|\epsilon_t|-\frac{1-2\tau+2\tau^2}{\tau(1-\tau)}\mathbf{1}_{n_t}\right)
\end{aligned}$$

Besides, $\mathbb{E}\left[(\epsilon_t-\xi_1\mathbf{1}_{n_t})\left(|\epsilon_t|-\frac{1-2\tau+2\tau^2}{\tau(1-\tau)}\mathbf{1}_{n_t}\right)^\top\right]=\frac{(1-2\tau)^3}{\tau^2(1-\tau)^2}I_{n_t}$. Thus, we have

$$\operatorname{Cov}(\mathbb{I}_1+\mathbb{I}_2,\mathbb{I}_3)=-\frac{\left(\sum_{l=1}^{t-1}n_l\right)n_t(1-2\tau)^2}{\left(\sum_{l=1}^{t}n_l\right)^2}\Sigma_{t-1}+\frac{n_t^2(1-2\tau)^4}{\left(\sum_{l=1}^{t}n_l\right)^2\tau^2(1-\tau)^2}(X_t^{*\top}X_t^*)^{-1} \tag{45}$$

Combining (39), (40), (41) and (44), (45), we finally obtain that

$$\Sigma_t=\frac{n_t^2 g(\tau)}{\left(\sum_{l=1}^{t}n_l\right)^2}(X_t^{*\top}X_t^*)^{-1}+\frac{\left(\sum_{l=1}^{t-1}n_l\right)^2+n_t^2(1-2\tau)^2+2\left(\sum_{l=1}^{t-1}n_l\right)n_t(1-2\tau)^4}{\left(\sum_{l=1}^{t}n_l\right)^2}\Sigma_{t-1}$$

where $g(\tau)=\frac{1-2\tau+2\tau^2+(1-2\tau)^2\left(1-2\tau-2\tau^2+8\tau^3-4\tau^4\right)-2(1-2\tau)^4}{\tau^2(1-\tau)^2}$.

6.4 Proof of Theorem 4

By (25), we have

$$\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}-\boldsymbol{\beta}=\left[\prod_{l=1}^{t}(1-r_l)\right](\overline{\boldsymbol{\beta}}_0^{\mathrm{QR}}-\boldsymbol{\beta})-\sum_{l=1}^{t}r_l\left[\prod_{m=l+1}^{t}(1-r_m)\right](\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}}) \tag{46}$$

where $r_t=\frac{n_t}{\sum_{l=1}^{t}n_l}$. Next, we separately analyze the two terms in the right-hand side of (46). Firstly, by using $\left[\prod_{l=1}^{t}(1-r_l)\right]=O\left(\frac{1}{t}\right)$, we have

$$\left\|\left[\prod_{l=1}^{t}(1-r_l)\right](\overline{\boldsymbol{\beta}}_0^{\mathrm{QR}}-\boldsymbol{\beta})\right\|_2=\left[\prod_{l=1}^{t}(1-r_l)\right]\left\|\overline{\boldsymbol{\beta}}_0^{\mathrm{QR}}-\boldsymbol{\beta}\right\|_2=O_p\left(\frac{1}{t}\right)$$

Secondly, as analyzed before, $(\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}}-\boldsymbol{\beta})$, $l=1,2,\ldots$ forms a martingale difference sequence, thus to control the norm of the second term, we need to control $\left\|r_l\left[\prod_{m=l+1}^{t}(1-r_m)\right](\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}})\right\|_2$ and $\sum_{l=1}^{t}\mathbb{E}\left\|r_l\left[\prod_{m=l+1}^{t}(1-r_m)\right](\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}})\right\|_2$.

By utilizing (37a), (37b), (37c), (38), we have

$$\begin{aligned}&\left\|r_l\left[\prod_{m=l+1}^{t}(1-r_m)\right](\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}})\right\|_2\\&=r_l\left[\prod_{m=l+1}^{t}(1-r_m)\right]\left\|(X_l^{*\top}X_l^*)^{-1}X_l^{*\top}(\xi_1\boldsymbol{v}_l+\xi_2\sqrt{\boldsymbol{v}_l}\circ\boldsymbol{u}_l-\xi_1\widetilde{\boldsymbol{v}}_l)\right\|_2\\&\le\left(\lambda_{\min}^{-1}\sqrt{\lambda_{\max}}/\sqrt{n_l}\right)r_l\left[\prod_{m=l+1}^{t}(1-r_m)\right]\left[\|\xi_1(\boldsymbol{v}_l-\widetilde{\boldsymbol{v}}_l)\|_2+\|\xi_2\sqrt{\boldsymbol{v}_l}\circ\boldsymbol{u}_l\|_2\right]=O_p\left(\frac{1}{t}\right)\end{aligned}$$

Further, we can obtain $\sum_{l=1}^{t}\mathbb{E}\left\|r_l\left[\prod_{m=l+1}^{t}(1-r_m)\right](\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}})\right\|_2=O\left(\frac{1}{t}\right)$. Then by applying the Pinelis-Bernstein inequality (see Proposition A.3 in [TY14]), we have

$$\left\|\sum_{l=1}^{t}r_l\left[\prod_{m=l+1}^{t}(1-r_m)\right](\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}_l^{\mathrm{QR}})\right\|=O_p\left(\frac{1}{t}\right)$$

and thus $\left\|\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}} - \boldsymbol{\beta}\right\|_2 = O_p\left(\frac{1}{t}\right)$. That is, the `Online-QR` estimator $\overline{\boldsymbol{\beta}}_t^{\mathrm{QR}}$ enjoys a linear convergence rate of $O\left(\frac{1}{t}\right)$.

6.5 Proof of Theorem 5

For the regret of `Online-QR`, we have

$$
\begin{aligned}
R_t(\boldsymbol{\omega}) &= \sum_{l=1}^{t} \frac{1}{n_l}\left[\sum_{i=1}^{n_l} \rho_\tau(y_{il} - {x_{il}^*}^\top \overline{\boldsymbol{\beta}}_l^{\mathrm{QR}}) - \sum_{i=1}^{n_l} \rho_\tau(y_{il} - {x_{il}^*}^\top \boldsymbol{\omega})\right] \\
&\leq \sum_{l=1}^{t} \frac{1}{n_l} \sum_{i=1}^{n_l} \rho_\tau({x_{il}^*}^\top \boldsymbol{\omega} - {x_{il}^*}^\top \overline{\boldsymbol{\beta}}_l^{\mathrm{QR}}) \qquad (47) \\
&\leq \max(\tau, 1-\tau) \sum_{l=1}^{t} \frac{1}{n_l} \sum_{i=1}^{n_l} \left|{x_{il}^*}^\top \boldsymbol{\omega} - {x_{il}^*}^\top \overline{\boldsymbol{\beta}}_l^{\mathrm{QR}}\right| \\
&\leq \max(\tau, 1-\tau) \frac{\sqrt{2}\max_l (n_t/n_l)}{n_t} \sqrt{\sum_{l=1}^{t} \left\|X_l^*(\boldsymbol{\omega} - \overline{\boldsymbol{\beta}}_l^{\mathrm{QR}})\right\|_2^2} \qquad (48) \\
&\leq \max(\tau, 1-\tau) \frac{\sqrt{2}\max_l (n_t/n_l)}{n_t} \sqrt{\sum_{l=1}^{t} \lambda_{\max} n_l \|\boldsymbol{\omega} - \overline{\boldsymbol{\beta}}_l^{\mathrm{QR}}\|_2^2} = O_p(\sqrt{t})
\end{aligned}
$$

In (47), we utilize the additive inequality of the check loss function (see Lemma 1 in [FLL23]), and in (48), we use the inequality between the l_1-norm and l_2-norm of a vector and the assumption $\max_l n_l / \min n_l = O(1)$.

7 Conclusion

In this work, we proposed an efficient online algorithm, `Online-QR`, for QR estimation in stream data, and further extended it to the `Multiple-QR` model. Our proposal is motivated by the Gibbs sampler for Bayesian QR and the iteration in the SA method for GLMM.

We converted the non-smooth check loss optimization to a normal linear regression estimation problem, and greatly improved the local processing speed because solving the converted problem only requires computing a least squares estimator. It is worth noting that the proposed `Online-QR` algorithm can be applied beyond the context of stream data. For example, when analyzing static big data stored distributedly over a network, standard divide-and-conquer techniques require the existence of a central node to perform the aggregation. However, in modern networks, this type of centralized computing is often a weakness to cyber attacks and decentralized computing is more preferred. Under such a situation, our `Online-QR` algorithm may serve as an efficient way to perform QR analysis when coupled with a smart way to traverse the network.

References

[ACFV06] Joshua Angrist, Victor Chernozhukov, and Iván Fernández-Val. Quantile regression under misspecification, with an application to the us wage structure. *Econometrica*, 74(2):539–563, 2006.

[AGPS09] David E Allen, Paul Gerrans, Robert Powell, and Abhay Kumar Singh. Quantile regression: its application in investment analysis. *Jassa*, (4):7–12, 2009.

[Alh16] Rahim Alhamzawi. Bayesian analysis of composite quantile regression. *Statistics in Biosciences*, 8(2):358–373, 2016.

[Ama13] Xavier Amatriain. Mining large streams of user data for personalized recommendations. *ACM SIGKDD Explorations Newsletter*, 14(2):37–48, 2013.

[AP12] Paul Alagidede and Theodore Panagiotidis. Stock returns and inflation: Evidence from quantile regressions. *Economics Letters*, 117(1):283–286, 2012.

[AWYZ21] Mingyao Ai, Fei Wang, Jun Yu, and Huiming Zhang. Optimal subsampling for large-scale quantile regression. *Journal of Complexity*, 62:101512, 2021.

[Bad13] Ihsan Ullah Badshah. Quantile regression analysis of the asymmetric return-volatility relation. *Journal of Futures Markets*, 33(3):235–265, 2013.

[Bah66] R Raj Bahadur. A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580, 1966.

[BK22] Bharti and Ashish Kumar. Exploring herding behaviour in indian equity market during covid-19 pandemic: impact of volatility and government response. *Millennial Asia*, 13(3):513–531, 2022.

[BR60] Raghu Raj Bahadur and R Ranga Rao. On deviations of the sample mean. *The Annals of Mathematical Statistics*, 31(4):1015–1027, 1960.

[BR02] Bernard Bercu and Alain Rouault. Sharp large deviations for the ornstein–uhlenbeck process. *Theory of Probability & Its Applications*, 46(1):1–19, 2002.

[BSX19] Aleksandr Beknazaryan, Hailin Sang, and Yimin Xiao. Cramér type moderate deviations for random fields. *Journal of Applied Probability*, 56(1):223–245, 2019.

[Buc98] Moshe Buchinsky. The dynamics of changes in the female wage distribution in the usa: a quantile regression approach. *Journal of applied econometrics*, 13(1):1–30, 1998.

[CBO21] Nicolò Cesa-Bianchi and Francesco Orabona. Online learning algorithms. *Annual review of statistics and its application*, 8(1):165–190, 2021.

[CFV11] Victor Chernozhukov and Iván Fernández-Val. Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies*, 78(2):559–589, 2011.

[CH06] Victor Chernozhukov and Christian Hansen. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525, 2006.

[Cha08] Sheng-Mao Chang. *A stationary stochastic approximation algorithm for estimation in the glmm.* North Carolina State University, 2008.

[CLMY20] Xi Chen, Weidong Liu, Xiaojun Mao, and Zhuoyi Yang. Distributed high-dimensional regression under a quantile loss function. *Journal of Machine Learning Research*, 21(182):1–43, 2020.

[CLZ19] Xi Chen, Weidong Liu, and Yichen Zhang. Quantile regression under memory constraint. *The Annals of Statistics*, 47(6):3244–3273, 2019.

[CMVW16] Michael Crouch, Andrew McGregor, Gregory Valiant, and David P Woodruff. Stochastic streams: Sample complexity vs. space complexity. In *24th Annual European Symposium on Algorithms (ESA 2016)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2016.

[Cra38] Harald Cramér. Sur un nouveau theoreme-limite de la theorie des probabilities. *Actualites Scientifiques et Industrielles*, 736:5–23, 1938.

[CZ20] Lanjue Chen and Yong Zhou. Quantile regression in big data: A divide and conquer based strategy. *Computational Statistics & Data Analysis*, 144:106892, 2020.

[Dag89] JS Dagpunar. An easily implemented generalised inverse Gaussian generator. *Communications in Statistics-Simulation and Computation*, 18(2):703–710, 1989.

[DZ98] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer Berlin, Heidelberg, 2nd edition, 1998.

[EI97] Peter Englund and Yannis M Ioannides. House price dynamics: an international empirical perspective. *Journal of housing economics*, 6(2):119–136, 1997.

[FFL09] Sergio Firpo, Nicole M Fortin, and Thomas Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

[FGL15] Xiequan Fan, Ion Grama, and Quansheng Liu. Sharp large deviation results for sums of independent random variables. *Science China Mathematics*, 58:1939–1958, 2015.

[FGLS20] Xiequan Fan, Ion Grama, Quansheng Liu, and Qi-Man Shao. Self-normalized cramér type moderate deviations for stationary sequences and applications. *Stochastic Processes and their Applications*, 130(8):5124–5148, 2020.

[FHM21] Xiequan Fan, Haijuan Hu, and Xiaohui Ma. Cramér moderate deviations for the elephant random walk. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(2):023402, 2021.

[FHX24] Xiequan Fan, Haijuan Hu, and Lihu Xu. Normalized and self-normalized Cramér-type moderate deviations for the Euler-Maruyama scheme for the SDE. *Science China Mathematics*, pages 1–16, 2024.

[FKM13] Bernd Fitzenberger, Roger Koenker, and José AF Machado. *Economic applications of quantile regression*. Springer Science & Business Media, 2013.

[FKMM22] Bernd Fitzenberger, Roger Koenker, José Machado, and Blaise Melly. Economic applications of quantile regression 2.0. *Empirical Economics*, 62(1):1–6, 2022.

[FLL23] Ye Fan, Jr-Shin Li, and Nan Lin. Residual projection for quantile regression in vertically partitioned big data. *Data Mining and Knowledge Discovery*, 37(2):710–735, 2023.

[GN17] Edward L Glaeser and Charles G Nathanson. An extrapolative model of house price dynamics. *Journal of Financial Economics*, 126(1):147–170, 2017.

[GSR13] Joao Gama, Raquel Sebastiao, and Pedro Pereira Rodrigues. On evaluating stream learning algorithms. *Machine learning*, 90:317–346, 2013.

[HPTZ22] Xuming He, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. Scalable estimation and inference for censored quantile regression process. *The Annals of Statistics*, 50(5):2899–2924, 2022.

[HPTZ23] Xuming He, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. Smoothed quantile regression with large-scale inference. *Journal of Econometrics*, 232(2):367–388, 2023.

[JLZ12] Yonggang Ji, Nan Lin, and Baoxue Zhang. Model selection in binary and tobit quantile regression using the gibbs sampler. *Computational Statistics & Data Analysis*, 56(4):827–839, 2012.

[Jou06] Cyrille Joutard. Sharp large deviations in nonparametric estimation. *Nonparametric Statistics*, 18(3):293–306, 2006.

[KBJ78] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[KK11] Hideo Kozumi and Genya Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation*, 81(11):1565–1578, 2011.

[Li24] Chris Junchi Li. A note on concentration inequality for vector-valued martingales with weak exponential-type tails. *Available at SSRN: `https://ssrn.com/abstract=4934923` or `http://dx.doi.org/10.2139/ssrn.4934923`*, 2024.

[LKLC03] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pages 2–11, 2003.

[LLSS24] Sokbae Lee, Yuan Liao, Myung Hwan Seo, and Youngki Shin. Fast inference for quantile regression with tens of millions of observations. *Journal of Econometrics*, page 105673, 2024.

[LS20] Lan Luo and Peter X-K Song. Renewable estimation and incremental inference in generalized linear models with streaming data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):69–97, 2020.

[LS23] Lan Luo and Peter X-K Song. Multivariate online regression analysis with heterogeneous streaming data. *Canadian Journal of Statistics*, 51(1):111–133, 2023.

[MCX10] Yu Miao, Ying-Xia Chen, and Shou-Fang Xu. Asymptotic properties of the deviation between order statistics and p-quantile. *Communications in Statistics—Theory and Methods*, 40(1):8–14, 2010.

[MLW+20] Wenlong Mou, Chris Junchi Li, Martin J Wainwright, Peter L Bartlett, and Michael I Jordan. On linear stochastic approximation: Fine-grained polyak-ruppert and non-asymptotic concentration. In *Conference on Learning Theory*, pages 2947–2997. PMLR, 2020.

[MP08] Lingjie Ma and Larry Pohlman. Return forecasts and optimal portfolio construction: a quantile regression approach. *The European Journal of Finance*, 14(5):409–425, 2008.

[MPTZ24] Rebeka Man, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. A unified algorithm for penalized convolution smoothed quantile regression. *Journal of Computational and Graphical Statistics*, 33(2):625–637, 2024.

[MZ23] Haeseong Moon and Wen-Xin Zhou. High-dimensional composite quantile regression: Optimal statistical guarantees and fast algorithms. *Electronic Journal of Statistics*, 17(2):2067–2119, 2023.

[Pet75] VV Petrov. *Sums of Independent Random Variables*. Springer-Verlag, Berlin, 1975.

[PK97] Stephen Portnoy and Roger Koenker. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4):279–300, 1997.

[Rei74] R-D Reiss. On the accuracy of the normal approximation for quantiles. *The Annals of Probability*, pages 741–744, 1974.

[Saa08] Jani Saastamoinen. Quantile regression analysis of dispersion of stock returns-evidence of herding? 2008.

[Ser09] Robert J Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 2009.

[Sta09] John Stachurski. *Economic dynamics: theory and computation*. MIT Press, 2009.

[TBZ22] Kean Ming Tan, Heather Battey, and Wen-Xin Zhou. Communication-constrained distributed quantile regression with optimal statistical guarantees. *Journal of machine learning research*, 23(272):1–61, 2022.

[TWZ22] Kean Ming Tan, Lan Wang, and Wen-Xin Zhou. High-dimensional quantile regression: Convolution smoothing and concave regularization. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):205–233, 2022.

[TY14] Pierre Tarres and Yuan Yao. Online learning as stochastic approximation of regularization paths: Optimality and almost-sure convergence. *IEEE Transactions on Information Theory*, 60(9):5716–5735, 2014.

[VCC19] Stanislav Volgushev, Shih-Kang Chao, and Guang Cheng. Distributed inference for quantile regression processes. *The Annals of Statistics*, 47(3):1634–1662, 2019.

[vWB22] Wessel N van Wieringen and Harald Binder. Sequential learning of regression models by penalized estimation. *Journal of Computational and Graphical Statistics*, 31(3):877–886, 2022.

[WWL22] Kangning Wang, Hongwei Wang, and Shaomin Li. Renewable quantile regression for streaming datasets. *Knowledge-Based Systems*, 235:107675, 2022.

[WWSH17] Jonathan Wang, Kesheng Wu, Alex Sim, and Seongwook Hwangbo. Accurate signal timing from high frequency streaming data. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 4852–4854. IEEE, 2017.

[WYW+23] Jiawei Wen, Songshan Yang, Christina Dan Wang, Yifan Jiang, and Runze Li. Feature-splitting algorithms for ultrahigh dimensional quantile regression. *Journal of Econometrics*, page 105426, 2023.

[XM11] Shoufang Xu and Yu Miao. Limit behaviors of the deviation between the sample quantiles and the quantile. *Filomat*, 25(2):197–206, 2011.

[YL17] Liqun Yu and Nan Lin. Admm for penalized quantile regression in big data. *International Statistical Review*, 85(3):494–518, 2017.

[YLW17] Liqun Yu, Nan Lin, and Lan Wang. A parallel algorithm for large-scale nonconvex penalized quantile regression. *Journal of Computational and Graphical Statistics*, 26(4):935–939, 2017.

[YM01] Keming Yu and Rana A Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.

[YMM14] Jiyan Yang, Xiangrui Meng, and Michael W Mahoney. Quantile regression for large-scale applications. *SIAM Journal on Scientific Computing*, 36(5):S78–S110, 2014.

[YNH18] Xiaorong Yang, Naveen Naidu Narisetty, and Xuming He. A new approach to censored quantile regression estimation. *Journal of Computational and Graphical Statistics*, 27(2):417–425, 2018.

[YS07] Keming Yu and Julian Stander. Bayesian analysis of a tobit quantile regression model. *Journal of Econometrics*, 137(1):260–276, 2007.

[YS18] Bicheng Ying and Ali H Sayed. Performance limits of stochastic sub-gradient learning, part i: Single agent case. *Signal Processing*, 144:271–282, 2018.

[YWH16] Yunwen Yang, Huixia Judy Wang, and Xuming He. Posterior inference in bayesian quantile regression with asymmetric Laplace likelihood. *International Statistical Review*, 84(3):327–344, 2016.

[YZ05] Keming Yu and Jin Zhang. A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics—Theory and Methods*, 34(9-10):1867–1879, 2005.

[ZDW13] Yuchen Zhang, John C Duchi, and Martin J Wainwright. Communication-efficient algorithms for statistical optimization. *The Journal of Machine Learning Research*, 14(1):3321–3363, 2013.

[ZW07] Qi Zhang and Wei Wang. A fast algorithm for approximate quantiles in high speed data streams. In *19th International Conference on Scientific and Statistical Database Management (SSDBM 2007)*, pages 29–29. IEEE, 2007.

[ZY08] Hui Zou and Ming Yuan. Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36(3):1108–1126, 2008.

A Experimental setup

We conducted an experiment to assess the performance of the `Online-QR` algorithm in estimating regression coefficients from streaming data. For illustrative purposes, we set the quantile level $\tau = 0.5$. The experiment was structured as follows:

- **Data Generation:** We generated synthetic data consisting of $N = 100,000$ observations over $T = 200$ time blocks. Each time block contained $n = \frac{N}{T}$ observations, with predictor variables $\mathbf{X}$ sampled from a $p = 50$ dimensional multivariate normal distribution $\mathcal{N}(\mathbf{0}_p, \mathbf{V})$, where $\mathbf{V}_{jk} = 0.5^{|j-k|}$.

- **Regression Model:** The response variable $\mathbf{y}$ was generated according to the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta} = (-1)^j$ for $j = 1, 2, \ldots, p$ represents the true regression coefficients, and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, 1)$ denotes the random error.

- **Algorithm Implementation:** We implemented the `Online-QR` algorithm to estimate the regression coefficients $\boldsymbol{\beta}$. The algorithm processed each time block in an online fashion, updating the estimate $\overline{\boldsymbol{\beta}}^{\mathrm{QR}}$ using QR decomposition at each step.

- **Evaluation Metrics:** We evaluated the algorithm's performance using two metrics:

    1. **Absolute Estimation Error (AE):** Defined as the sum of absolute differences between the estimated and true regression coefficients across all dimensions.
    2. **ℓ_1-norm Distance:** The sum of absolute differences between the final estimated coefficients $\overline{\boldsymbol{\beta}}_T^{\mathrm{QR}}$ and the true coefficients $\boldsymbol{\beta}$.

The R code used for data generation, algorithm implementation, and metric computation is shown below. Each run of the code produces varying outputs due to the random nature of data generation, ensuring robustness and generalizability of the experiment results.

```
# Load necessary library
library(MASS)

# Set parameters
N <- 100000
T <- 200
n <- N / T
p <- 50

# Generate the covariance matrix V
V <- matrix(0, nrow = p, ncol = p)
for (j in 1:p) {
  for (k in 1:p) {
    V[j, k] <- 0.5^abs(j - k)
  }
}

# Generate the covariates from the multivariate normal distribution
```

```r
X <- MASS::mvrnorm(n = N, mu = rep(0, p), Sigma = V)

# Generate the regression coefficients
beta <- (-1)^(1:p)

# Generate the random error epsilon
epsilon <- rnorm(N)

# Calculate y using the linear model
y <- X %*% beta + epsilon

# Split data into T blocks
X_blocks <- array(NA, dim = c(n, p, T))
y_blocks <- matrix(NA, nrow = n, ncol = T)

for (t in 1:T) {
  start_idx <- (t - 1) * n + 1
  end_idx <- t * n
  X_blocks[,,t] <- X[start_idx:end_idx, ]
  y_blocks[,t] <- y[start_idx:end_idx]
}

# Initialize beta_QR_0
bar_beta_QR <- rep(0, p)

# Initialize AE
AE <- 0

# Algorithm for QR estimation in stream data
for (t in 1:T) {
  # (a) Store (X_t, y_t) and bar_beta_QR(t-1)
  X_t <- X_blocks[,,t]
  y_t <- y_blocks[,t]

  # (b) Compute tilde_v_t and tilde_beta_QR_t
  tilde_v_t <- rep(0, n)
  tilde_beta_QR_t <- rep(0, p)

  for (i in 1:n) {
    hat_delta_it2 <- (1/8) * (y_t[i] - X_t[i,] %*% bar_beta_QR)^2
    tilde_v_it <- hat_delta_it2 + 0.5
    tilde_v_t[i] <- tilde_v_it
  }

  X_t_star <- cbind(1, X_t)  # Add intercept to X_t
  y_t_star <- y_t
```

```
  # Compute tilde_beta_QR_t using QR solution
  tilde_beta_QR_t <- solve(t(X_t_star) %*% X_t_star) %*% t(X_t_star) %*% y_t_star

  # (c) Update bar_beta_QR_t
  if (t == 1) {
    bar_beta_QR <- tilde_beta_QR_t[-1]  # Initialize bar_beta_QR_1 without intercept
  } else {
    bar_beta_QR <- (1/t) * tilde_beta_QR_t[-1] + ((t-1)/t) * bar_beta_QR
  }

  # Compute AE for this iteration and accumulate
  AE <- AE + sum(abs(bar_beta_QR - beta))
}

# Compute the L1-norm distance between bar_beta_QR and beta
L1_distance <- sum(abs(bar_beta_QR - beta))

# Print results
cat("Final Estimate bar_beta_QR_T:\n")
print(c(0, bar_beta_QR))  # Add intercept back for final estimate

cat("Absolute Estimation Error (AE):\n")
print(AE)

cat("L1-norm distance between bar_beta_QR and beta:\n")
print(L1_distance)
```

This experimental setup allowed us to assess the effectiveness and accuracy of the `Online-QR` algorithm in handling streaming data and estimating regression coefficients under varying data volumes and dimensions.

A.1 Experimental results

To evaluate the performance of the `Online-QR` algorithm, we ran the experiment using synthetic data generated as described earlier. Below are the results from a sample run:

```
> cat("Absolute Estimation Error (AE):\n")
Absolute Estimation Error (AE):
> print(AE)
[1] 61.03646

> cat("L1-norm distance between bar_beta_QR and beta:\n")
L1-norm distance between bar_beta_QR and beta:
> print(L1_distance)
[1] 0.1681378
```

These metrics quantify the accuracy of our algorithm in estimating the regression coefficients under the experimental conditions.