Renewable composite quantile method and algorithm for nonparametric models with streaming data

Yan Chen · Shuixin Fang · Lu Lin

Received: 8 October 2022 / Accepted: 25 October 2023 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
We are interested in renewable estimations and algorithms for nonparametric models with streaming data. In our method, the nonparametric function of interest is expressed through a functional depending on a weight function and a conditional distribution function (CDF). The CDF is estimated by renewable kernel estimations together with function interpolations, based on which we propose the method of renewable weighted composite quantile regression (WCQR). Then, by fully utilizing the model structure, we obtain new selectors for the weight function, such that the WCQR can achieve asymptotic unbiasedness when estimating specific functions in the model. We also propose practical bandwidth selectors for streaming data and find the optimal weight function by minimizing the asymptotic variance. The asymptotical results show that our estimator is almost equivalent to the oracle estimator obtained from the entire data together. Besides, our method also enjoys adaptiveness to error distributions, robustness to outliers, and efficiency in both estimation and computation. Simulation studies and real data analyses further confirm our theoretical findings.

Keywords Renewable algorithm · Streaming data · Composite quantile regression · Nonparametric regression · Polynomial interpolation

1 Introduction
1.1 Problem setup and challenges
In this paper, we are interested in nonparametric regression problems for massive data taking the form of streaming data. Specifically, the considered nonparametric model is that

\[ Y = m(X) + \sigma(X)\varepsilon, \]  

(1)

where \( Y \) and \( X \) are supposed respectively to be scalar response variable and covariate for simplicity; \( \varepsilon \) is the random error independent of \( X \), and the distribution of \( \varepsilon \) is unknown and satisfies \( \mathbb{E}[\varepsilon] = 0 \) and \( \text{Var}[\varepsilon] = 1 \); \( m : \mathbb{R} \to \mathbb{R} \) and \( \sigma : \mathbb{R} \to [0, \infty) \) are both unknown functions. In this paper, the model (1) is said to be symmetric if the probability density function (PDF) of \( \varepsilon \) is symmetric around 0, and otherwise the model (1) is said to be asymmetric. The considered streaming data consist of a series of cross-sectional data chunks \( D_t = \{(X_{tj}, Y_{tj}) : 1 \leq j \leq n_t\} \) for \( t = 1, 2, \ldots \), where all \( (X_{tj}, Y_{tj}) \) are independent and identically distributed (i.i.d.) observations of \( (X, Y) \). In our setting, the data chunks \( D_1, D_2, \ldots \) are not available simultaneously, but arrive sequentially one after another.

As we know, most conventional statistic algorithms are designed under the premise that the full data can be fitted on the computer memory simultaneously. However, such a premise is no longer true for streaming data mentioned above. To deal with streaming data, the online-updating (or renewable) algorithms are widely considered. For example, at the time \( t \), one has obtained a summary statistic \( T_t \) of the historical data \( \cup_{s \leq t} D_s \). Then as the new data chunk \( D_{t+1} \) arrives, \( T_t \) is updated to \( T_{t+1} \) by incremental computation without accessing the historical raw data, i.e., \( T_{t+1} = R(T_t; D_{t+1}) \) with \( R \) a function independent of \( \cup_{s \leq t} D_s \). When modified
into the above renewable form, the statistics may lose desirable statistical properties, which brings new challenges in designing statistical algorithms for online-updating.

The first challenge arises from the data partitioning. As we know, nonparametric methods inevitably suffer from estimation bias. The bias cannot be reduced by simply averaging the local estimators from each data chunk, which essentially prevents the renewable estimator from achieving the standard statistical convergence rate. Hence when designing algorithms for streaming data, it is crucial to sufficiently reduce the estimation bias.

The second challenge lies in the potentially poor quality of streaming data. Outliers and fat-tailed features are more likely to hide in these massive raw data. And even worse, it is quite hard to detect or address them, because the relevant procedures usually involve reusing the historical statistics. Thus it has a significant value for renewable algorithms that the obtained estimator is robust to outliers and is adaptive to fat-tailed features.

The third challenge is caused from the exploding data size. The streaming data source usually generates extremely large amounts of data in a short period of time. To deal with such a rapid data stream, the updating algorithm should be implemented efficiently.

1.2 Existing works and motivations

There have been many works developed for streaming data. Usually, the existing online-updating methods can be classified into the following categories. In some restrictive cases, the estimator has a closed-form expression and then its exact value can be obtained by some recursive updating operations, see, e.g., Schifano et al. (2016), Bucak and Günsel (2009), Nion and Sidiropoulos (2009), etc. However, it is more often the case that the expression of an estimator is complex, then iterative algorithms of online-updating are often used to approximate the value of the estimator, see, e.g., Robbins and Monro (1951), Toulis et al. (2014), Moroshko et al. (2015), Chen et al. (2019), etc. Additionally, several online cumulative frameworks are proposed for likelihood, estimating equations and so on, see, e.g., Luo and Song (2020), Lin et al. (2020), Wang et al. (2022), etc. And there are also some works based on the deep learning techniques, see, Ashfahani and Pratama (2019), Das et al. (2019), Pratama et al. (2019), to name a few. In this case, the related iterative algorithms of online-updating are more complicated.

When the statistic has a closed-form expression, it is easy to enjoy exactly the same statistical properties as that of the oracle estimator obtained by the offline methods together with the full dataset. However, such a result often relies on such a closed-form expression, which is unavailable for most robust estimators including the quantile estimators. Without the closed-form expression, the differentiability condition of the objective functions is required to achieve the oracle property, see, e.g., Luo and Song (2020) and Lin et al. (2020). However, it is often the case where the robust objective functions (e.g., quantile based objective functions) are not differentiable.

In this paper, we will address the above issues via a new approach for streaming data. By the new proposal, the renewable algorithm does not rely on the closed-form expression of a statistic, but the obtained estimator still achieves the standard statistical convergence rate. Meanwhile, we also deal with the issues in the three challenges aforementioned, pursuing the robustness to outliers and the adaptiveness to various error distributions, and computational efficiency.

1.3 Contributions and article frame

In this paper, a renewable composite quantile method and the corresponding algorithm are proposed to estimate the nonparametric functions in the model (1) with streaming data.

Inspired by L-estimation (see, e.g., Koenker and Portnoy 1987; Portnoy and Koenker 1989; Boente and Fraiman 1994), we express the nonparametric function through a functional, instead of a closed-form expression. Here the functional takes the form of an integral depending on a weight function and a conditional distribution function (CDF) of $Y$. Then the renewable estimation is attained by two steps:

1. **Numerical Approximation**: The CDF in the functional is approximated by function interpolations. Then the nonparametric function can be approximately expressed by a finite number of function values of the CDF.

2. **Statistical Approximation**: The aforementioned function values of CDF are estimated by kernel estimators, which have closed-form expressions and can be exactly obtained through recursive updating algorithms.

By combining the above numerical and statistical approximations, we finally propose our renewable weighted composite quantile regression (WCQR) method for streaming data.

By the renewable WCQR, the functions $m(\cdot)$ and $\sigma(\cdot)$ can be estimated by correctly selecting the weight function. Specifically speaking, we fully use the structure of the model (1) and obtain new selection criterions for the weight functions, under which the renewable method can estimate $m(\cdot)$ and $\sigma(\cdot)$ asymptotically unbiased. Further, we deduce the asymptotic distributions of the proposed estimators. Based on the theoretical conclusions, a practical bandwidth selector is proposed for the online-updating estimator, and the optimal weight function is also obtained by minimizing the asymptotic variance under the constraint of the above selection criterions. Finally, our theoretical findings are demonstrated by simulation studies and real data analyses.
Compared with the competitors, our method has the following main virtues:

1) *Oracle comparability.* Through numerical approximations, our WCQR estimator is assembled from some renewable statistics exactly obtained via online-updating. Thanks to this, not only the algorithm gets rid of any restriction on the chunk size or chunk number of the streaming data, but also the obtained estimator enjoys almost the same asymptotic properties as that of the oracle estimator obtained on the full data set. Such a virtue is not enjoyed by many estimation methods for massive data, e.g., Schifano et al. (2016), Shang and Cheng (2017), Volgushev et al. (2019), Wang et al. (2022), etc.

2) *Robustness.* Benefit from robust feature of the quantile estimation, when the model has a symmetric error distribution, our regression method enjoys the robustness compared with the common methods such as ordinary least squares (Fan 1993; Ruppert et al. 1997).

3) *Model adaptiveness.* With model-based weight functions, our CQR method is adaptive to models with symmetric or asymmetric error distribution. Different from existing methods (Sun et al. 2013; Lin et al. 2019; Jiang et al. 2016), the selection criteria of the weight functions are directly established on the structure of the nonparametric models instead of the information of the errors. Thus in our method, the weight selection does not rely on any pilot estimations for the error distributions. Furthermore, the model adaptiveness can be preserved in case of streaming data.

4) *Estimation efficiency.* Under the above weight criterions, we find the optimal weight function, under which our renewable estimator enjoys minimized asymptotic variance in estimating \( m(x) \) and \( \sigma(x) \). Our estimators inherit the efficiency of CQR-typed estimators (Zou and Yuan 2008; Kai et al. 2010; Sun et al. 2013), which can be superior to the ordinary least squares (Fan 1993; Ruppert et al. 1997) for nonnormal error distributions.

5) *Computational efficiency.* Thanks to the closed-form expression of the kernel estimators, our algorithm is quite simple in updating procedures without need for solving any optimization problems or nonlinear equations. Compared with optimization-based methods, e.g., Kai et al. (2010), Sun et al. (2013), Wang et al. (2021), Wang et al. (2022), our method is quite computationally efficient. This feature is particularly desirable for rapid data stream.

The paper is then organized in the following way. In Sect. 2, we give some preliminaries about the existing composite quantile estimations and the L-estimation. In Sect. 3, the main idea of our methodologies is introduced in detail, including the computation of the renewable WCQR estimator, the selection of the weight functions, and some specific estimators and detailed renewable algorithms for estimating \( m(x) \) and \( \sigma(x) \). In Sect. 4, the asymptotic properties of the proposed method are established; the selector of online-updating bandwidth is proposed and the optimal selection of the weight function is discussed. Section 5 contains comprehensive simulation studies and real data analyses to further demonstrate the desirable performance of the proposed estimators and algorithms. The detailed algorithm, lemmas, and technical proofs are deferred to the supplementary material.

### 1.4 Notations

For a random variable \( Z \), denote by \( f_Z(\cdot) \) and \( F_Z(\cdot) \) the probability density function and the distribution function of \( Z \), respectively, and denote by \( Q_{Y|x}(\tau) \) the conditional \( 100\tau \% \) quantile of \( Y \) given \( X = x \in \mathbb{R} \) and \( \tau \in (0, 1) \), i.e., \( Q_{Y|x}(\tau) = \inf \{ t : \mathbb{P}(Y \leq t|X = x) \geq \tau \} \). Let \( F_{Y|x}(\cdot) \) be the conditional distribution function (CDF) of \( Y \) given \( X = x \), i.e., \( F_{Y|x}(y) = \mathbb{P}(Y \leq y|X = x) \), and \( f_{Y|x}(\cdot) \) be the conditional PDF of \( Y \) given \( X = x \).

For a, \( b \in \mathbb{R} \), denote \( a \land b = \min\{a, b\} \). For indexes \( i \) and \( j \), denote by \( \delta_{ij} \) the kronecker symbol, i.e., \( \delta_{ij} = 1 \) if \( i = j \) and \( \delta_{ij} = 0 \) if \( i \neq j \). Let \( I_d(\cdot) \) be the identity function, i.e., \( I_d(y) = y \), and let \( I(\cdot) \) stand for the indicative function, i.e., if the proposition \( \mathcal{P} \) is true, \( I(\mathcal{P}) = 1 \), and otherwise, \( I(\mathcal{P}) = 0 \). For a function \( f : \mathbb{R} \to \mathbb{R} \), denote by \( \text{Supp} \{ f(\cdot) \} \) the support set of \( f \); for \( I_0 \subset \mathbb{R} \) and \( 1 \leq p < \infty \), denote \( \| f \|_{p, I_0} = (\int_{I_0} |f(y)|^p \, dy)^{1/p} \) and \( \| f \|_{\infty, I_0} = \sup_{y \in I_0} |f(y)| \). For \( \tau \in [0, 1] \), denote the check function \( \rho_\tau(\cdot) \) by \( \rho_\tau(u) = u(\tau - I(u \leq 0)) \) for \( u \in \mathbb{R} \).

For a finite point set \( G \subset \mathbb{R} \), denote \( \Delta(G) = \max_{y_i \in G} \min_{y_j \in G \setminus \{y_i\}} |y_i - y_j| \), i.e., the maximum spacing between any two adjacent points in \( G \); in \#\,\# the number of elements in \( G \), and denote by \( \min(G) \) the minimum (resp. maximum) of \( G \). For two sets \( G_1 \) and \( G_2 \), denote \( G_1 \setminus G_2 = \{ y : y \in G_1, y \notin G_2 \} \). For a set \( E \subset \mathbb{R} \) and a function \( f : \mathbb{R} \to \mathbb{R} \), write \( f(E) = \{ f(x) : x \in E \} \); denote \( I(E) = [\inf E, \sup E] \) the closed interval generated by \( E \).

### 2 Preliminary

In this section, we will briefly review some fundamental works, from which we obtain inspirations for our method.
2.1 Composite quantile regression and L-estimation

To estimate the regression function \( m(x) \), the well-known composite quantile regression (CQR) estimator (see, e.g., Koenker and Zhao 1994) takes the form

\[
\hat{m}_{cqr}(x) = \frac{1}{q} \sum_{i=1}^{q} \hat{Q}_{Y|x}(\tau_i),
\]

where \( \hat{Q}_{Y|x}(\tau_i) \) are some consistent estimators of the quantile \( Q_{Y|x}(\tau_i) \), and \( \tau_i \) are quantile levels chosen as \( \tau_i = i/(q+1) \). When the PDF of \( \varepsilon \) is symmetric around 0, one can expect \( m(x) = 1/q \sum_{i=1}^{q} \hat{Q}_{Y|x}(\tau_i) \) implying that \( \hat{m}_{cqr}(x) \) is an asymptotic unbiased estimator of \( m(x) \). However, when the error is general, the aforementioned equality is not necessarily true and the naïve CQR estimator may suffer from non-negligible bias. To address this issue, Sun et al. (2013) extended the CQR estimator to the WCQR estimator in from of

\[
\hat{m}_{wcqr}(x) = \sum_{i=1}^{q} \omega_i \hat{Q}_{Y|x}(\tau_i),
\]

where \( \omega_i \) are weights selected to satisfy \( \sum_{i=1}^{q} \omega_i = 1 \) and \( \sum_{i=1}^{q} \omega_i F^{-1}_x(\tau_i) = 0 \), such that the equality \( m(x) = \sum_{i=1}^{q} \omega_i Q_{Y|x}(\tau_i) \) holds, consequently, the non-negligible bias of \( \hat{m}_{wcqr}(x) \) is eliminated.

The CQR and WCQR estimators can be expressed as L-estimators (see, e.g., Gutenbrunner and Jurečková 1992; Koenker and Zhao 1994) taking the form

\[
\hat{m}_{L}(x; v) = \int_{[0,1]} \hat{Q}_{Y|x}(\tau) \cdot dv(\tau),
\]

with \( v \) a measure on \([0, 1]\). To cover the CQR (resp. WCQR) estimators, the measure of L-estimators can be particularly chosen as \( v = 1/q \sum_{i=1}^{q} \delta_{\tau_i} \) (resp. \( v = \sum_{i=1}^{q} \omega_i \delta_{\tau_i} \)) with \( \delta_{\tau_i} \) a unit mass on \( \tau_i \). Additionally, when the measure \( v \) is well chosen based on the error distribution in (1), the L-estimator can efficiently (or robustly) estimate \( m(x) \) and \( \sigma(x) \) (see, e.g., Serfling 1980; Koenker and Portnoy 1987; Portnoy and Koenker 1989; Koenker 2005, etc.).

In this paper, we still call \( \hat{m}_{L}(x; v) \) the WCQR estimator to highlight the weighted composite of quantile estimators.

2.2 Local polynomial interpolation

We will briefly review the main idea of local polynomial interpolation (LPI), which is a fundamental tool for numerical approximations and will be applied in our method. For more details about the interpolation techniques and their applications to stochastic computing, readers may refer to Gautschi (2012), Sauer (2011), Burden et al. (2015), Zhao et al. (2006), Zhao et al. (2014), Fu et al. (2017), etc.

Suppose that \((y_1, f_1), \cdots, (y_q, f_q)\) are given data point generated from an unknown objective function \( f : \mathbb{R} \rightarrow \mathbb{R} \), i.e., \( f_i = f(y_i) \) for \( i = 1, \cdots, q \). Denote the node set \( G = \{y_i\}_{i=1}^{q} \) and the elements \( y_i \) are called the interpolation nodes.

The so-called LPI aims at finding a piecewise polynomial function passing through all the given data points of \( f(\cdot) \). To achieve this, for \( 1 \leq l \leq q \) and \( y \in \mathbb{R} \), we denote by \( N_l(y, G) \) the set consisting of the \( l \) nearest interpolation nodes around \( y \), or strictly speaking, \( N_l(y, G) = G_0 \) with \( G_0 \) the unique set satisfying:

a) \#\(G_0 \equiv l, G_0 \subseteq G\);

b) for any \( y' \in G_0 \) and \( y'' \in G \setminus G_0 \), it holds that \(|y - y'| \leq |y - y''| \), and whenever the equality holds, \( y' < y'' \).

Then we define the \( l \)th-degree LPI basis functions as

\[
L(y, y_i; l, G) = \ell_l(y) \mathbb{I}(y_i \in N_{l+1}(y, G))
\]

for \( y_i \in G \) with \( \ell_l(\cdot) \) the \( l \)th-degree Lagrange interpolating polynomials defined as

\[
\ell_l(y) = \prod_{y_j \in N_{l+1}(y, G) \setminus \{y_i\}} \left( y-y_i \right) \prod_{y_j \in N_{l+1}(y, G) \setminus \{y_i\}} \left( y-y_j \right)
\]

Using the basis functions, the \( l \)th-degree LPI function is constructed as

\[
I_l f(y) = \sum_{y_i \in G} f(y_i) L(y, y_i; l, G).
\]
where \( c \) is a number depending on \( y \) and lying in 
\([\min \hat{N}_{I+1}(y), \max \hat{N}_{I+1}(y)]\) with \( \hat{N}_{I+1}(y) = N_{I+1}(y, G) \) \( \cup \{y\} \).

### 3 Methodology

As stated in Introduction, the WCQR estimators enjoy robustness and efficiency for non-normal error discussion (Zou and Yuan 2008; Kai et al. 2010; Sun et al. 2013), and thus they are quite desirable for massive streaming data. However, the quantile estimators in (3) have no closed-form expressions, which make it difficult to construct a renewable WCQR estimator with its performance as well as the oracle counterparts obtained on the full dataset.

In this section, we will address the above issue. Specifically, we aim at proposing a renewable WCQR estimation and algorithm to estimate \( m(x) \) and \( \sigma(x) \) for \( x \in I_1 \) with \( I_1 \) a bounded interval on \( \mathbb{R} \). Here the proposed algorithm should get rid of any restrictions on the batch number \( T \) and the batch sizes \( n_t \) of the streaming data, and the obtained estimator should enjoy desirable statistical properties almost the same with the oracle WCQR estimator computed on the full data set.

#### 3.1 Renewable WCQR estimation

We start form a special case of (4), where the measure \( \nu \) exists a density function \( J(\cdot) \) on \([0, 1]\), and the WCQR estimator \( \hat{m}_L(x; \nu) \) can be expressed as

\[
\hat{r}(x; J) = \int_{[0, 1]} J(\tau) \hat{Q}_{Y|x}(\tau) \, d\tau. \tag{7}
\]

By selecting appropriate \( J(\cdot) \), the estimator \( \hat{r}(x; J) \) is able to estimate a kind of parameters that can be expressed as

\[
r(x; J) = \int_{[0, 1]} J(\tau) Q_{Y|x}(\tau) \, d\tau. \tag{8}
\]

To obtain \( \hat{r}(x; J) \), the conventional approach is to minimize an \( L_1 \)-norm loss function characterized by the check function \( \rho_1(u) = u(\tau - I(u \leq 0)) \). However, solving such a non-smooth minimization problem is quite difficult when the data are of the form of streaming data sets. To address this issue, we manage to avoid estimating the quantiles in (3). Thus we introduce the substitution: \( y = \hat{Q}_{Y|x}(\tau) \), i.e., \( \tau = \hat{Y}_{Y|x}(y) \) and rewrite (7) into

\[
\hat{r}(x; J) = \int_{\mathbb{R}} y J(\hat{Y}_{Y|x}(y)) \, d\hat{F}_{Y|x}(y), \tag{9}
\]

where \( \hat{F}_{Y|x}(y) \) is an estimator of the CDF \( F_{Y|x}(y) \).

Now the key problem is to obtain a renewable estimation of \( F_{Y|x}(\cdot) \). To this end, we approximate \( F_{Y|x}(\cdot) \) by its LPI function given by

\[
\tilde{I}_l F_{Y|x}(y) = \sum_{y_i \in G_x} F_{Y|x}(y_i) L(y, y_i; l, G_x), \tag{10}
\]

where \( G_x = \{y_i\}_{i=1}^q \) is the set of interpolation nodes chosen for \( x \in I_1 \) and \( L(\cdot, y_i; l, G_x) \) are LPI basis functions defined in (5). Then the unknown values \( F_{Y|x}(y_i) \) can be estimated by their empirical analogues obtained from the streaming data sets \( D_1, D_2, \ldots \), i.e.,

\[
\hat{F}_{Y|x,t}(y_i) = \frac{\hat{S}_{Y|x,t}(y_i)}{\hat{f}_{X,t}(x)} \text{ for } y_i \in G_x, \tag{11}
\]

where \( \hat{f}_{X,t}(x) \) and \( \hat{S}_{Y|x,t}(y_i) \) are renewable statistics obtained by

\[
\hat{f}_{X,t}(x) = \frac{N_t - 1}{N_t} \hat{f}_{X,t-1}(x) + \frac{1}{N_t} \sum_{j=1}^{n_t} K_{h_l}(X_{ij} - x), \tag{12}
\]

\[
\hat{S}_{Y|x,t}(y_i) = \frac{N_{t-1}}{N_t} \hat{S}_{Y|x,t-1}(y_i) + \frac{1}{N_t} \sum_{j=1}^{n_t} I(Y_{ij} < y_i) K_{h_l}(X_{ij} - x) \tag{13}
\]

with initial values \( N_0 = \hat{f}_{X,0}(x) = \hat{S}_{Y|x,0}(y_i) = 0 \) and the bandwidths \( h_l > 0 \). Based on (10), we plug in the estimators \( \hat{F}_{Y|x,t}(y_i) \) of \( F_{Y|x}(y_i) \) and obtain the interpolated empirical CDF as follows

\[
\tilde{I}_l \hat{F}_{Y|x,t}(y) = \sum_{y_i \in G_x} \hat{F}_{Y|x,t}(y_i) L(y, y_i; l, G_x). \tag{14}
\]

By (9) with \( \tilde{I}_l \hat{F}_{Y|x,t}(y) \) in place of \( \hat{F}_{Y|x,t}(y) \), we can obtain the renewable WCQR estimator as

\[
\tilde{r}(x; J) = \int_{\mathbb{R}} y J \left( \tilde{I}_l \hat{F}_{Y|x,t}(y) \right) d\tilde{I}_l \hat{F}_{Y|x,t}(y). \tag{15}
\]

Since the expression of \( \tilde{I}_l \hat{F}_{Y|x,t}(\cdot) \) is known, the integral in (15) can be accurately approximated by numerical integrations, e.g., the well-known trapezoidal rule, Simpson rule and Romberg integration, etc. (see, e.g., Sect. 3 of Gautschi 2012).

By applying LPI on the \( x \)-axis, we can extend the pointwise estimator \( \tilde{r}(x; J) \) to estimate the function \( r(\cdot; J) \) on

\[\text{ Springer}\]
the entire interval \( x \in I_s \). Specifically, we can approximate \( r(\cdot; J) \) by its LPI function, i.e.,

\[
\mathcal{I}_J r(x; J) = \sum_{x_i \in G_s} r(x_i; J) L(x, x_i; I, G_s),
\]

(16)

where \( G_s = \{ x_i \}_{i=1}^q \) is a set of grid points introduced on the interval \( I_s \), and typically \( x_i \) can be chosen as equal-spaced grid points on \( I_s \). With \( r(x_i; J) \) estimated by \( \tilde{r}_i(x_i; J) \), we can obtain the renewable interpolated WCQR estimator as

\[
\mathcal{I}_J \tilde{r}_i(x; J) = \sum_{x_i \in G} \tilde{r}_i(x_i; J) L(x, x_i; I, G_s)
\]

(17)

for \( x \in I_s \). In Sect. 3.3, we will see that \( \mathcal{I}_J \tilde{r}_i(\cdot; J) \) plays a role in selecting the weight function \( J(\cdot) \) for streaming data.

For the restriction on the weight function in (15), we have the following remark.

**Remark 3.1** By Lemma 2.1, we can conclude that

\[
\| \mathcal{I}_J F_{Y|x} - F_{Y|x} \|_{\infty, I(G_s)} \leq \| F_{Y|x}^{[l+1]} \|_{\infty, I(G_s)} | \Delta (G_s)|^{l+1}
\]

with \( I(G_s) = [\min G_s, \max G_s] \). Thus if \( F_{Y|x}(\cdot) \) is sufficiently smooth and the nodes are dense enough, the error caused from LPI can be negligible on \( I(G_s) \). However, when \( y \) moves away from the interval \( I(G_s) \), the LPI approximation cannot guarantee its accuracy. Fortunately, we can select appropriate \( J(\cdot) \) to suppress the error of LPI when \( y \) lies outside \( I(G_s) \). Specifically, for the renewable WCQR estimator \( \tilde{r}_i(x; J) \), we can select \( J(\cdot) \) satisfying

\[
\text{Supp} \{ J(\cdot) \} \subset \mathcal{I}_J \tilde{F}_{Y|x}^{-1} (I(G_s)).
\]

(18)

And for the interpolated WCQR estimator \( \mathcal{I}_J \tilde{r}_i(\cdot; J) \), we can select \( J(\cdot) \) satisfying (18) for all \( x \in G_s \) with \( G_s \) satisfying \( I(G_s) \supset I_s \). Given \( I(G_s) \) wide enough, the above restrictions are mild in robust estimations, because \( \text{Supp} \{ J(\cdot) \} \subset (0, 1) \) is a natural condition to guarantee the robustness of \( \tilde{r}(x; J) \) (see, e.g., Sect. 8.1.3 of Serfling 1980).

In the remainder of this section, we mainly discuss the selection of the weight function \( J(\cdot) \), which plays a key role in reducing the estimation bias and variance of our renewable WCQR estimator.

### 3.2 Model-based weight selections

To estimate \( m(x) \) (resp. \( \sigma(x) \)) by renewable WCQR estimation, we should select appropriate weight functions \( J(\cdot) \) to establish the equality \( r(x; J) = m(x) \) (resp. \( \sigma(x) \)). To this end, recalling (8) and using the relation \( \mathcal{Q}_{Y|x}(\tau) = m(x) + \sigma(x) F_{\tau}^{-1}(\tau) \) obtained from the model (1), we can deduce

\[
r(x; J) = m(x) \int_{[0, 1]} J(\tau) d\tau
\]

+ \( \sigma(x) \int_{[0, 1]} J(\tau) F_{\tau}^{-1}(\tau) d\tau, \)

(19)

which yields the conditions for estimating \( m(x) \):

\[
C_{m1}: \int_{[0, 1]} J(\tau) d\tau = 1,
\]

(20)

and the conditions for estimating \( \sigma(x) \):

\[
C_{m2}: \int_{[0, 1]} J(\tau) F_{\tau}^{-1}(\tau) d\tau = 0,
\]

(21)

Among the above conditions, \( C_{m1} \) and \( C_{m2} \) can be easily satisfied. However, \( C_{m2} \) and \( C_{m2} \) are related to the unknown quantile function \( F_{\tau}^{-1}(\cdot) \). Unless the error distribution is known to be symmetric, it is quite difficult to choose \( J(\cdot) \) in an renewable manner; see the following remark.

**Remark 3.2** In the existing works, e.g., Sun et al. (2013), Lin et al. (2019), Jiang et al. (2016), the condition \( C_{m2} \) is fulfilled empirically by replacing the unknown function \( F_{\tau}^{-1}(\cdot) \) with its estimator \( \widehat{F}_{\tau}^{-1}(\cdot) \). Here \( \widehat{F}_{\tau}^{-1}(\cdot) \) is the sample quantile function of the “pseudo” samples \( \hat{\epsilon}_{ti} = (Y_{ti} - \hat{m}(X_{ti})) / \hat{\sigma}(X_{ti}) \), where \( \hat{m}(\cdot) \) and \( \hat{\sigma}(\cdot) \) are some pilot estimators of \( m(\cdot) \) and \( \sigma(\cdot) \). The generation of pseudo samples involves reusing the historical raw data and requires sophisticated computations, which are hardly to be implemented for streaming data.

To avoid the problem mentioned above, we fully use the structure of the Model (1) and obtain the following important lemma, which gives an alternative way to fulfill the conditions \( C_{m2} \) and \( C_{m2} \).

**Lemma 3.1** Let \( W : \mathbb{R} \to \mathbb{R} \) be a function satisfying \( \mathbb{E}[W(X)\sigma(X)] > 0 \), then the following results hold:

(i) if \( C_{m1} \) holds, then

\[
\mathbb{E}[W(X)\sigma(X)] \int_{[0, 1]} J(\tau) F_{\tau}^{-1}(\tau) d\tau
\]

\[
= \mathbb{E}[W(X)(r(X; J) - Y)]
\]

\[
\mathbb{E}[W(X)(r(X; J) - Y)]
\]

\[
\mathbb{E}[W(X)(r(X; J) - Y)]
\]

\[
\mathbb{E}[W(X)(r(X; J) - Y)]
\]
(ii) if \( C_{\sigma_1} \) holds, then

\[
\begin{align*}
\mathbb{E} \left[ W(X)r^2(X; J) \right] & = \mathbb{E} \left[ W(X) \right]^2 \left( \int_{[0,1]} J(\tau)F_{\epsilon}^{-1}(\tau) \, d\tau \right)^2 \\
& = \mathbb{E} \left[ W(X) \left( Y^2 - m^2(X) \right) \right].
\end{align*}
\]

Lemma 3.1 suggests the following alternative conditions on the weight function:

\[
\begin{align*}
C_{m2}' : & \quad \mathbb{E} \left[ W(X) r(X; J) \right] = \mathbb{E} \left[ W(X) Y \right], \quad (22) \\
C_{\sigma 2}' : & \quad \mathbb{E} \left[ W(X) r^2(X; J) \right] = \mathbb{E} \left[ W(X) \left( Y^2 - m^2(X) \right) \right]. \quad (23)
\end{align*}
\]

Lemma 3.1 also shows that \((C_{m1}, C_{m2}) \Leftrightarrow (C_{m1}', C_{m2}')\) and \((C_{\sigma 1}, C_{\sigma 2}) \Leftrightarrow (C_{\sigma 1}', C_{\sigma 2}')\), i.e., the condition \(C_{m2}\) (resp. \(C_{\sigma 2}\)) can be equivalently replaced with \(C_{m2}'\) (resp. \(C_{\sigma 2}'\)). The following remark shows the advantage of \(C_{m2}'\) over \(C_{m2}\).

**Remark 3.3** As stated in Remark 3.2, to fulfill \(C_{m2}\), the existing methods in Sun et al. (2013), Lin et al. (2019), Jiang et al. (2016) rely on the estimation of the inverse of the CDF of \(\epsilon\), which is unavailable directly from the sample set of \((X, Y)\). Instead of directly related to the distribution of the error, the condition \(C_{m2}'\) only relies on the two expectations in (22), which can be directly estimated by the sample of \((X, Y)\), and the estimation has the convergence rate of parametric estimation. Moreover, in Sect. 3.3.1, we will see that \(C_{m2}'\) can be easily expressed in a renewable form.

Although the condition \(C_{\sigma 2}'\) in (23) relies on the unknown regression function \(m(\cdot)\), it will be shown in Sect. 3.3.2 that the unknown \(m(\cdot)\) does not bring any essential difficulty to our renewable estimation.

### 3.3 Specific renewable WCQR estimators and algorithms

By constructing specific \(J(\cdot)\) satisfying the conditions in the previous subsection, we can obtain specific WCQR estimators for \(m(x)\) or \(\sigma(x)\). To present some examples of \(J(\cdot)\), throughout this subsection, we assume the function \(W(\cdot)\) in (22) and (23) satisfies

\[
\begin{align*}
\mathbb{E} [W(X) \sigma(X)] > 0, & \quad \text{Supp} \{W(\cdot)\} \subset I_*.
\end{align*}
\]

where the first condition is required by Lemma 3.1, and the second condition is used to avoid estimating \(r(x; J)\) for \(x\) outside of \(I_*\). The simplest \(W(\cdot)\) satisfying (24) is that

\[
W(x) = I(x \in I_*) \quad \text{for} \quad x \in \mathbb{R}.
\]

### 3.3.1 Estimators for the conditional mean

We first consider a simple case, where the model (1) is symmetric. For symmetric models, \((C_{m1}, C_{m2})\) can be easily fulfilled whenever \(J(\cdot)\) is normalized and symmetric around 1/2. A representative example is the \(\alpha\)-trimmed weight function:

\[
J_{m,0.5}(\tau) = 0.5L_{\alpha}(\tau) + 0.5U_{\alpha}(\tau),
\]

where

\[
\begin{align*}
L_{\alpha}(\tau) & = (0.5 - \alpha)^{-1} I(\alpha < \tau \leq 0.5), \\
U_{\alpha}(\tau) & = (0.5 - \alpha)^{-1} I(0.5 < \tau \leq 1 - \alpha)
\end{align*}
\]

with \(\alpha \in (0, 0.5)\) selected according to Remark 3.1. By the weight function \(J_{m,0.5}(\cdot)\), the WCQR estimator \(\tilde{r}_I(x; J_{m,0.5})\) given by (15) is a renewable version of the naive local \(\alpha\)-trimmed mean (NTM) (see, e.g., Bednar and Watt 1984; Boente and Fraiman 1994). In the following, we will call \(\tilde{r}_I(x; J_{m,0.5})\) the renewable NTM or the NTM if there is no confusion.

The following remark shows that the NTM is robust to outliers, and only adaptive to symmetric but not to asymmetric error distributions.

**Remark 3.4** The prototype of \(\tilde{r}_I(x; J_{m,0.5})\) is the \(\alpha\)-trimmed mean, which has been singled out by several prominent authors as the quintessential robust estimator of location (see, e.g., Bickel and Lehmann 1975; Stigler 1977; Koenker 2005). And we can expect that \(\tilde{r}_I(x; J_{m,0.5})\) enjoys robustness comparable to the \(\alpha\)-trimmed mean. However, since \(\tilde{r}_I(x; J_{m,0.5})\) actually estimate the location \(r(x; J_{m,0.5})\) rather than the conditional mean \(m(x)\), the estimation consistency deeply relies on the symmetry of the error to guarantee the conditions \((C_{m1}, C_{m2})\) and the equality \(r(x; J_{m,0.5}) = m(x)\). When the error distribution is asymmetric, the estimator \(\tilde{r}_I(x; J_{m,0.5})\) will suffer from a non-negligible bias caused from \(r(x; J_{m,0.5}) \neq m(x)\).

To address the issue mentioned in Remark 3.4, we should modify the \(\alpha\)-trimmed weight function \(J_{m,0.5}(\cdot)\), such that...
with the initial value $\hat{E}_{Wy,0} = 0$. Here $I_t\hat{r}_t(\cdot; L_\alpha)$ and $I_t\hat{r}_t(\cdot; U_\alpha)$ are the interpolated WCQR estimator given in (17), and $I_t f_X(\cdot)$ is the interpolated empirical PDF given by

$$I_t f_X(\cdot) = \sum_{x_i \in G_s} \hat{f}_{x,t}(x_i) L(\cdot, x_i; l, G_s)$$ (33)

with $\hat{f}_{x,t}(x_i)$ the renewable statistics obtained by (12) and $G_s$ the node set introduced in (16).

Compared with the estimation error, the approximation error of LPI can be ignored, implying that $\hat{w}_t$ is a consistent estimator of $w$. Thus we known from the Slutsky’s Lemma that the estimator $\hat{r}_t(x; J_m, \hat{w}_t)$ has the same asymptotic distribution as that of $\hat{r}_t(x; J_m, w)$.

In the following remark, we show the pros and cons of the BCTM compared with the NTM.

**Remark 3.6** Compared with the NTM, the first advantage of the BCTM is that the estimation consistency is based on the structure of the model, instead of the symmetry of the error. Thus the BCTM is adaptive to symmetric or asymmetric error distributions. Although this adaptiveness comes with a little costs in robustness, when estimating the parameter $\omega$, it is easy to check that the estimators $E_{Wy,t}$ and $E_{W2,t}$ have bounded influence functions, implying that both of them are robust to outliers. However, the estimator $E_{Wy,t}$ is somewhat non-robust. Fortunately, the unknown $E_{Wy,t}$ is a scalar parameter rather than a general function, thus the estimator $E_{Wy,t}$ can achieve $\sqrt{N_t}$-consistency, which is faster than the optimal convergence rate of nonparametric regression estimation. Hence the final nonparametric estimator $\hat{r}_t(x; J_m, \hat{w}_t)$ is less susceptible to the weak robustness of $E_{Wy,t}$. This point will be further demonstrated in our numerical studies in Sect. 5.

In Algorithm 1, we present the detailed computational procedures to obtain the renewable BCTM $\hat{r}_t(x; J_m, \hat{w}_t)$ for the estimation of $m(x)$ with $x$ belonging to some grid points in $I_1$. We note that on Line 21 of Algorithm 1, the following trick is used to accelerate the computation:

$$\hat{r}_t(x; J_m, \hat{w}_t) = \hat{w}_t \tilde{r}(x; L_\alpha) + (1 - \hat{w}_t) \tilde{r}(x; U_\alpha)$$ (34)

for $x \in G_s$, where $\tilde{r}(x; L_\alpha)$ and $\tilde{r}(x; U_\alpha)$ are obtained on Lines 14 and 15 of Algorithm 1 to obtain $\hat{w}_t$. The validity of (34) follows from (15) and (27).

**3.3.2 Estimators for the conditional variance**

To estimate $\sigma(x)$, the first condition $C_{\alpha1}$ can be fulfilled by taking $J(\cdot)$ antisymmetry around 1/2. Then parallel to
Algorithm 1: Renewable BCTM for estimating \([m(x_i)]_{i=1}^n\)

Input: Set of grid points \(G_x = \{x_i\}_{i=1}^n\), kernel function \(K(\cdot)\), node sets \(G_x\) for \(x_i \in G_x\), LP degree \(l\)

1: Set initial values \(N_0 = \hat{E}_{WY,0} = \hat{f}_{XY,0} (x_i) = \hat{S}_{Y|x_i}(0, y_j) = 0\) for \(y_j \in G_y\) and \(x_i \in G_x\)
2: for \(t = 1, 2, \ldots\) do
3: Obtain the \(t\)-th data chunk \(D_t = \{X_{ij}, Y_{ij}\}_{j=1}^n\)
4: Select the \(t\)-th bandwidth \(h_t\)
5: \(N_t = N_{t-1} + n_t\)
6: \(\hat{E}_{WY,t} = \hat{N}_{t-1}/N_t \hat{E}_{WY,t-1} + 1/N_t \sum_{j=1}^n W(X_{ij}) Y_{ij}\)
7: for \(x_i \in G_x\) do
8: \(\hat{f}_{X|Y_t} (x_i) = \frac{N_t}{N_{t-1}/N_t \hat{f}_{X|Y_{t-1}} (x_i) + 1/N_t \sum_{j=1}^n K_h (X_{ij} - x_i)\), where \(K_h(\cdot) = 1/h K(\cdot/h)\)
9: \(\hat{S}_{Y|X_t} (y_j) = \frac{N_t}{N_{t-1}/N_t \hat{S}_{Y|X_{t-1}} (y_j) + 1/N_t \sum_{j=1}^n I (Y_{ij} < y_j) K_h (X_{ij} - x_i)\) for \(y_j \in G_y\)
10: end for
11: if the estimators for \([m(x_i)]_{x_i \in G_x}\) are needed then
12: Define the function \(\hat{I}_t \hat{f}_{X,t} (\cdot) = \sum_{x_i \in G_x} \hat{f}_{X|Y_t} (x_i) L (\cdot, x_i, l, G_x)\)
13: Define the function \(\hat{I}_t \hat{S}_{Y|X_t} (\cdot) = \sum_{y_j \in G_y} \hat{S}_{Y|X_t} (y_j) / \hat{f}_{X|Y_t} (x_i) L (\cdot, y_j, l, G_x)\)
14: for \(J \in \{l_a, u_a\}\) do
15: \(\hat{t}_r (x_i; J) = \frac{\int_{J_a} y J (\hat{I}_t \hat{f}_{X,t} (x_i)) d \hat{S}_{Y|X_t} (y) / \hat{f}_{X|Y_t} (x_i)}{\int_{J_a} \hat{f}_{X|Y_t} (x_i) L (\cdot, l, G_x)}\)
16: Define the function \(\hat{I}_t \hat{t}_r (\cdot ; J) = \sum_{x_i \in G_x} \hat{t}_r (x_i; J) L (\cdot, x_i, l, G_x)\)
17: end for
18: \(\hat{E}_{WY,t} = \int_{J_a} W (x) \hat{I}_t \hat{t}_r (x; l_a, \bar{u}_a) \hat{I}_t \hat{f}_{X,t} (x) dx\)
19: \(\hat{E}_{WU,t} = \int_{J_a} W (x) \hat{I}_t \hat{f}_{X,t} (x; l_a, \bar{u}_a) \hat{I}_t \hat{f}_{X,t} (x) dx\)
20: \(\bar{w}_t = (\hat{E}_{WY,t} - \hat{E}_{WU,t}) / (\hat{E}_{WY,t} - \hat{E}_{WU,t})\)
21: \(\hat{r}_t (x_i; J_m, \bar{w}_t) = \bar{w}_t \hat{r}_t (x_i; \bar{u}_a) + (1 - \bar{w}_t) \hat{r}_t (x_i; l_a)\)
22: Output \([\hat{r}_t (x_i; J_m, \bar{w}_t)]_{x_i \in G_x}\) as the estimators for \([m(x_i)]_{x_i \in G_x}\)
23: end if
24: end for

From \((19)\) and \((35)\), we can see that

\[
J_{\sigma,1} (\tau) = (-L_\alpha (\tau) + U_\alpha (\tau)) .
\]  

(35)

The main issue is to obtain a renewable estimation of the unknown \(\theta\). Similar to the idea introduced in Sect. 3.3.1, we can estimate \(\theta\) by

\[
\hat{\theta}_t = \sqrt{\frac{\hat{E}_{WY^2,t} - \hat{E}_{W^2,t}}{\hat{E}_{W^2,t}}}.
\]

(39)

The initial value \(\hat{E}_{WY^2,0} = 0\). Here \(\hat{I}_t \hat{r}_t (\cdot ; J_{\sigma,1})\) and \(\hat{I}_t \hat{f}_{X,t} (\cdot)\) are the rescaled \(\alpha\)-trimmed standard deviations \(\hat{E}_{WY^2,t} \) and \(\hat{E}_{W^2,t}\), respectively.

Remark 3.7 The RTSD \(\hat{r}_t (x; J_{\sigma,\bar{\alpha}})\) contains more information about \(\sigma (x)\), since it can identify the constant \(c_0\), which is unambiguous by the RTSD \(\hat{r}_t (x; J_{\sigma,1})\). In another aspect, the RTSD enjoys desirable robustness comparable to the

In the following text, we call \(\hat{r}_t (x; J_{\sigma,1})\) the naive \(\alpha\)-trimmed standard deviation (NTSD), since it is modified from the NTM and is used to estimate the conditional standard derivation.

If our goal is to consistently estimate \(\sigma (x)\), we should rescale the weight function such that \(c_0 = 1\), or equivalently, \(C_{\sigma,2}\) holds. To this end, we extend \(J_{\sigma,1} (\cdot)\) into

\[
J_{\sigma,0} (\tau) = \theta J_{\sigma,1} (\tau),
\]

(37)

where \(\theta\) is a scale parameter selected to satisfy the alternative \(C_{\sigma,2}\), i.e.,

\[
\theta^2 = \frac{\mathbb{E} \left[ W(X) \left( Y^2 - m^2 (X) \right) \right]}{\mathbb{E} \left[ W(X) r^2 (X; J_{\sigma,1}) \right]} .
\]

(38)
classical $\alpha$-trimmed mean. While the RTSD depends on the plug-in estimator $\hat{E}_{W^2,t}$, and similar to the discussions in Remark 3.6, $\hat{E}_{W^2,t}$ is somewhat non-robust but enjoys the convergence rate of parametric estimation.

In the supplementary material, we present a complete algorithm to obtain renewable WCQR estimators for $m(x)$ and $\sigma(x)$ from asymmetric models.

In the following remark, we show the desirable computation efficiency of our algorithms in dealing with rapid data steams.

**Remark 3.8** Algorithm 1 mainly consists of two parts: the updating part (Lines 5–10) where the cumulative statistics are updated, and the estimation part (Lines 12–21) where the WCQR estimator is computed by calculating integrals. Notice that the updating part is relatively simple without solving any nonlinear equations. And at each updating step, the computations among each $x_i \in G_s$ can be implemented in parallel. Benefit from this, the updating part can be implemented fast enough to catch up with the rapid data stream. The estimation part seems to be computationally intensive. Fortunately, this part is implemented only when one needs the current value of the WCQR estimator. Thus it would not cause trouble in computation speed even when the data stream is rapid. The above desirable feature is also enjoyed by Algorithm A.1 in the supplementary material.

### 3.4 Extension to multivariate models

In the previous subsections, we consider the nonparametric model (1) with univariate just for clearly introducing our main idea. Actually, our methodology is also applicable to nonparametric models with multivariate. To show this, we consider a multivariate version of (1) as

$$Y = m(X) + \sigma(X) \epsilon,$$

where $X$ is a $d$-dimensional covariate, and $Y$ is a scalar response. We aim at estimating $m(x)$ and $\sigma(x)$ for $x = (x_1, \ldots, x_d)^T \in I^d_u$ with $I_u$ a bounded interval on $\mathbb{R}$. In this case, the streaming data chunks are denoted by $D_t = \{(X_{ij}, Y_{ij}) : 1 \leq j \leq n_t\}$ for $t = 1, 2, \ldots$, where $(X_{ij}, Y_{ij})$ are i.i.d. samples of $(X, Y)$. The $l$-th element of $X_{ij}$ is denoted as $X_{ij,l}$.

Notice that $Y$ is still a scalar response, and the derivations in the above subsections can be trivially extended to the model (43). For brevity, we only present the constructions of estimators for (43) as follows.

The renewable WCQR estimator for regression function $m(x)$ in model (43) is given by

$$\hat{r}(x; J) = \int_{\mathbb{R}} y J \left( \int_{\mathbb{R}} F_{Y|x,t}(y) \right) d\int_{\mathbb{R}} F_{Y|x,t}(y)$$

with

$$\int_{\mathbb{R}} \hat{F}_{Y|x,t}(y) \sum_{y_i \in G_x} \hat{F}_{Y|x,t}(y_i) L(y, y_i; l, G_x),$$

where $G_x = \{ y_i \}_{i=1}^{n_d} \subset \mathbb{R}$ is the set of interpolation nodes chosen for $x \in I^d_u$. $F_{Y|x,t}(\cdot)$ is the empirical CDF of $Y|X = x$ obtained by (11) - (13) with the notation $x$, $x$, $K_{ht}(X_{ij} - x)$ replaced by $X$, $X$, $\prod_{l=1}^{d} K_{ht}(X_{ij,l} - x_l)$, respectively.

To estimate the regression function $m(x)$ and variance function $\sigma(x)$, the weight function $J(\cdot)$ in $\hat{r}_t(x; J)$ can be respectively selected as $J_{m, \hat{w}_t}(\cdot)$ and $J_{\sigma, \hat{w}_t}(\cdot)$. Here the estimator $\hat{w}_t$ in $J_{m, \hat{w}_t}(\cdot)$ is given by (29) - (32) with the notation $I_s, x, X, L(X_{ij})$ being replaced by $I_s, x, X, \prod_{l=1}^{d} W(X_{ij,l})$, respectively. Similarly, the estimator $\hat{w}_t$ in $J_{\sigma, \hat{w}_t}(\cdot)$ is given by (39)–(42) with the same replacements.

### 4 Theoretical analysis and selections of bandwidth and weight

In this section, we will deduce the asymptotic distribution of the renewable WCQR estimator, and based on the proposed theoretical conclusion, we will conduct renewable bandwidth selectors and obtain the optimal weight functions. For simplicity, all the analyses in this section focus on the univariate model (1). The restriction on the dimensionality of $X$ is not essential, because the presented analysis methods and results can be easily extended to the multivariate model (43).

#### 4.1 Asymptotic distribution

For theoretical analysis of the estimator $\hat{r}_t(x; J)$, we introduce the following regularity conditions.

**Assumption 4.1** The functions $f_X(\cdot)$, $m(\cdot)$ and $\sigma(\cdot)$ have continuous derivatives up to order 4 on $I_u$. On a open set containing $\text{Supp} \{ J(\cdot) \}$, the function $F_t(\cdot)$ has continuous derivatives up to order max $\{ l + 1, 4 \}$, where $l$ is the degree of LP1 given in (10). The function $f_X(\cdot)$ admits a positive lower bounded on $I_u$.

**Assumption 4.2** The kernel function $K(u)$ is symmetric and compactly supported, and satisfies that $\int_{\mathbb{R}} K(u)du = 1$, $k_{4,1} = \int_{\mathbb{R}} u^4 K(u)du < \infty$ and $k_{0,2} = \int_{\mathbb{R}} K^2(u)du < \infty$.

**Assumption 4.3** The bandwidth $h_l$ satisfies that as $t \to \infty$,

$$h_l = o(1), \quad \left\{ \frac{1}{N_l} \sum_{s=1}^{n_s} h_s \right\}^{-1/2} \sum_{s=1}^{n_s} h_s^4 = o(1),$$

$$\frac{1}{N_t} \sum_{s=1}^{n_s} \frac{n_s}{N_t h_s} = o(1).$$

$$\frac{1}{N_t} \sum_{s=1}^{n_s} \frac{n_s}{N_t h_s} = o(1).$$
Assumption 4.4 The L-score function $J(\cdot)$ is piecewise continuously differentiable. There exist constants $0 < \tau < \tau' < 1$ such that $\text{Supp} \{ J(\cdot) \} \subset [\tau, \tau']$. The node set $G_x$ satisfies $[\tau, \tau'] \subset (\min F_{Y|x} (G_x), \max F_{Y|x} (G_x))$.

In Assumptions 4.1 and 4.2, all the conditions on $f_X(\cdot)$, $m(\cdot)$, $\sigma(\cdot)$ and $K(\cdot)$ are standard for nonparametric regressions, and the condition on $F_x(\cdot)$ is required by the LPI approximation to guarantee its accuracy. Among the conditions in Assumption 4.3, the first one in (44) is required for a vanishing estimation bias; the second one in (44) is used to simplify the bias terms in the asymptotic results, and it holds under the condition $h_1 = o(N_t^{-1/9})$ following from Jensen’s inequality; the last condition (45) is necessary for a vanishing estimation variance as $t \to \infty$, and it is fulfilled whenever $N_t \min_{1 \leq s \leq t} h_s \to \infty$. Assumption 4.4 is used to bound the remainder term in the asymptotic expansion of $\tilde{r}_t(x; J)$; the restrictions on $\text{Supp} \{ J(\cdot) \}$ and $G_x$ are also required by the LPI approximation and in practical applications, it can be fulfilled empirically; see Remark 3.1.

The following theorem gives the asymptotic properties of the interpolated empirical CDF $\tilde{F}_{y|x, t}(\cdot)$.

Theorem 4.1 Under Assumptions 4.1–4.3, it holds for $y \in \mathbb{R}$ that

$$\left\{ \frac{1}{N_t} \sum_{s=1}^{t} \frac{n_s}{N_t h_s} \right\}^{-1/2} \left\{ \tilde{I}_t F_{Y|x} (y) - F_{Y|x} (y) \right. \right.$$

$$- R_{\tilde{I}_t, F, x} (y) - \tilde{I}_t B_{F, x} (y) \left( \sum_{s=1}^{t} \frac{n_s h_s^2}{N_t} \right) \}
\xrightarrow{d} N \left( 0, \frac{Z^2}{\tilde{I}_t^2} C_{F, x} (y, y) \right),$$

where

$$R_{\tilde{I}_t, F, x} (y) = - \frac{F_{Y|x}^{(l+1)}(c)}{(l+1)!} \prod_{y \in N_t \cap (y; G_x)} (y - y_i),$$

$$\tilde{I}_t B_{F, x} (y) = \sum_{y \in G_x} B_{F, x} (y) \left( L(y, y_i; l, G_x) \right),$$

$$\tilde{I}_t^2 C_{F, x} (z_1, z_2) = \sum_{y, y_j \in G_x} L(z_1, y_i; l, G_x) C_{F, x} (y_i, y_j) \times L(z_2, y_j; l, G_x) C_{F, x} (y_i, y_j)$$

with

$$B_{F, x} (y) = - \frac{k_{x, 1}}{2} \left[ \frac{\partial^2 F_{Y|x} (y_i)}{\partial y^2} \right] + 2 \sigma_{x, 1} F_{Y|x} (y_i) \left[ \frac{f_{x} (y)}{f_{x} (x)} \right],$$

$$C_{F, x} (y_i, y_j) = \frac{k_{0, 2}}{f_{x} (x)} \left[ F_{Y|x} (y_i \wedge y_j) \right]$$

and $c \in \left[ \min \mathcal{N}_{H+1} (y), \max \mathcal{N}_{H+1} (y) \right]$ with $\mathcal{N}_{H+1} (y)$ given in Lemma 2.1.

For better understanding the convergence rate obtained by Theorem 4.1, we introduce the following decomposition representation:

$$\left\| \tilde{I}_t \tilde{F}_{Y|x, t} - F_{Y|x} \right\|_{2, l(G_x)} = O(|\Delta (G_x)|^{l+1}) + O \left( \sum_{s=1}^{t} \frac{n_s h_s^2}{N_t} \right) + O_P \left( \frac{1}{N_t} \sum_{s=1}^{t} \frac{n_s}{N_t h_s} \right).$$

The proof of (48) is given in the Supplementary Material. From the representation in (48) we can see that the error of our interpolated empirical CDF consists of two parts: the first part determined by $\Delta (G_x)$ is the numerical error caused from the LPI approximation, and the second part that depends on the bandwidths consists of statistical errors from the bias and variance of kernel estimations. Because the statistical errors have a convergence rate as the same as that of the oracle empirical CDF obtained on the imaginary full data set $\cup_{x \in D} \mathcal{D}_t$, and the numerical error is usually negligible compared with the statistical ones (see the following Remark 4.1), it can be expected that our estimator enjoys a performance almost as well as the oracle estimator.

Remark 4.1 In practical applications, the numerical error $O(|\Delta (G_x)|^{l+1})$ in (48) is usually much smaller than the remaining statistical errors. Actually, when the nodes in $G_x$ are uniformly spaced, the numerical error is of order $O((\#G_x)^{-l/2})$, which can be reduced significantly by increasing the number $\#G_x$ of nodes and the degrees $l$ of LPI, whenever the computation resources permit. Unlike the numerical one, the statistical errors deeply rely on the number $N_t$ of samples, and the convergence rate is relatively slow (no more than $O(N_t^{-2/5})$).

Based on Theorem 4.1, we have the following main result on the asymptotic property of the estimator $\tilde{r}_t(x; J)$.

Theorem 4.2 Under Assumptions 4.1–4.4, it holds that

$$\left\{ \frac{1}{N_t} \sum_{s=1}^{t} \frac{n_s}{N_t h_s} \right\}^{-1/2} \left\{ \tilde{r}_t(x; J) - r(x; J) \right.$$

$$- B_{\tilde{I}_t, m, x} \sum_{s=1}^{t} \frac{n_s h_s^2}{N_t} - R_{\tilde{I}_t, m, x} \}
\xrightarrow{d} N \left( 0, \Sigma_{\tilde{I}_t, m, x} \right),$$

 Springer
where
\[
B_{T_{x}, m, x} = - \int_{\mathbb{R}} J (F_{Y | x} (y)) I_{l} B_{F, x} (y) \, dy,
\]
\[
R_{T_{x}, m, x} = - \int_{\mathbb{R}} J (F_{Y | x} (y)) R_{T_{l}, F, x} (y) \, dy + O \left( \left| \Delta (G_{x}) \right|^{l+1} \right),
\]
\[
\Sigma_{T_{x}, m, x} = \int_{\mathbb{R}} I_{l} C_{I_{l}, x} (z_{1}, z_{2}) \, dz_{1} \, dz_{2}
\]

and \( I_{l} B_{F, x} (y), \ R_{T_{l}, F, x} (y) \) and \( I_{l} C_{I_{l}, x} (z_{1}, z_{2}) \) given in Theorem 4.1.

Theorem 4.2 shows that the LPI operator \( I_{l} \) has influence on the asymptotic bias and variance of \( \tilde{r}_{l} (x; J) \), while its influence is no more than the order of numerical errors. The following corollary states the details.

**Corollary 4.3** Under the conditions of Theorem 4.2 and with the notations defined in Theorem 4.2, it holds that
\[
B_{T_{x}, m, x} = B_{m, x} + O \left( \left| \Delta (G_{x}) \right|^{l+1} \right),
\]
\[
R_{T_{x}, m, x} = O \left( \left| \Delta (G_{x}) \right|^{l+1} \right),
\]
\[
\Sigma_{T_{x}, m, x} = \Sigma_{m, x} + O \left( \left| \Delta (G_{x}) \right|^{l+1} \right),
\]

where the dominant terms are
\[
B_{m, x} = - \int_{\mathbb{R}} J (F_{Y | x} (y)) B_{F, x} (y) \, dy,
\]
\[
\Sigma_{m, x} = \sigma^{2} \left( \int_{\mathbb{T}} \int_{\mathbb{T}} S_{J} (\tau_{1}, \tau_{2}) \, d\tau_{1} \, d\tau_{2} \right),
\]

with
\[
S_{J} (\tau_{1}, \tau_{2}) = \left( \frac{\tau_{1} \wedge \tau_{2} - \tau_{1} \tau_{2}}{\tau_{1} \tau_{2}} \right) J (\tau_{1}) J (\tau_{2}) f_{x} \left( F_{x}^{-1} (\tau_{1}) \right) f_{x} \left( F_{x}^{-1} (\tau_{2}) \right).
\]

Combining Theorem 4.2 and Corollary 4.3, we can find that the difference between \( \tilde{r}_{l} (x; J) \) and \( r (x; J) \) has a convergence rate as the same as the error given in (48), i.e.,
\[
\tilde{r}_{l} (x; J) - r (x; J) = O \left( \left| \Delta (G_{x}) \right|^{l+1} \right) + O_{p} \left( \frac{1}{N_{l}} \sum_{s=1}^{t} \frac{n_{s} h_{s}^{2}}{N_{s}} \right).
\]

Here, the first term is the numerical error caused from LPI, and the remaining two terms are statistical errors caused from kernel estimations. By the discussion in Remark 4.1, the numerical error is usually negligible compared with the statistical ones. Moreover, the convergence rates of the statistical errors are the same as that of the oracle estimator \( \tilde{r} (x; J) \) obtained on the imaginary full data set \( I_{l} \cup C_{1} \cup D_{1} \). Theoretically, our renewable WCQR estimator behaves almost as well as the oracle estimator.

Based on the relation (19) and the results in Lemma 3.1 and Theorem 4.2, we can immediately obtain the following corollary, which gives asymptotic properties of the estimators introduced in Sect. 3.3.

**Corollary 4.4** If \( J (\cdot) \) satisfies \((C_{m1}, C_{m2})\) or \((C_{m1}, C_{m2})\) (resp. \((C_{r1}, C_{r2})\) or \((C_{r1}, C_{r2})\)), then the result (49) in Theorem 4.2 holds with \( r (x; J) \) replaced with \( m (x) \) (resp. \( \sigma (x) \)).

### 4.2 Selection of bandwidths

For implementing bandwidth selection, we need to calculate the asymptotic mean square error (AMSE) and the asymptotic mean integrated squared error (AMISE) between \( \tilde{r}_{l} (x; J) \) and \( r (x; J) \). It follows from Theorem 4.2 and Corollary 4.3 that the asymptotic bias and variance of \( \tilde{r}_{l} (x; J) \) can be given by

\[
\text{Bias} \left( \tilde{r}_{l} (x; J) \right) = B_{m, x} = \sum_{s=1}^{t} \frac{n_{s} h_{s}^{2}}{N_{s}}
\]

\[
+ o \left( \sum_{s=1}^{t} \frac{n_{s} h_{s}^{2}}{N_{s}} \right) = O \left( |\Delta (G_{x})|^{l+1} \right),
\]

\[
\text{Var} \left( \tilde{r}_{l} (x; J) \right) = \frac{1}{N_{l}} \sum_{s=1}^{t} \frac{n_{s}}{N_{l} h_{s}} \Sigma_{m, x}
\]

\[
+ O \left( 1 + \frac{1}{N_{l}} \sum_{s=1}^{t} \frac{n_{s} h_{s}}{N_{s}} \right).
\]

By Remark 4.1, it is reasonably to assume the numerical error is negligible compared with the statistical ones, thus we can omit the terms \( O \left( |\Delta (G_{x})|^{l+1} \right) \) and other high order terms in the asymptotic bias and variance. Then the AMSE and AMISE can be respectively defined as

\[
\text{AMSE} \left( \tilde{r}_{l} (x; J) \right) = \mathcal{E}_{l} (x; H_{l}) \text{,}
\]

\[
\text{AMISE} \left( \tilde{r}_{l} (x; J) \right) = \int_{I_{s}} \mathcal{E}_{l} (x; H_{l}) \tilde{W} (x) \, dx,
\]

where \( \tilde{W} (\cdot) \) is a weight function defined on \( I_{s} \); \( H_{l} = (h_{1}, \cdots, h_{t})^{T} \) is a vector consisting of the used bandwidth components \( h_{s} \) up to the \( t \)-th data chunk and \( \mathcal{E}_{l} (x; \cdot) \) is an
error function defined by

$$E_t(x; H_t) = \left( \frac{n_s h_t^2}{N_T} \right)^2 \frac{1}{N_t} \sum_{s=1}^{t} \frac{n_s}{N_t} \Sigma_{m,x}$$

(52)

with the asymptotic bias \(B_{m,x}\) and the asymptotic variance \(\Sigma_{m,x}\) given in Corollary 4.3.

### 4.2.1 Theoretical optimal bandwidths

We first consider an ideal situation where the streaming data is finite with the terminal time \(T\) and the cumulative number \(N_T\) known throughout the updating procedures \(t = 1, \cdots, T\). In this case, the theoretical optimal variable bandwidths \(\tilde{H}_T^*(x) = (\tilde{h}_1^*(x), \cdots, \tilde{h}_T^*(x))^{\top}\) and the optimal constant bandwidth \(\tilde{H}_T^* = (\tilde{h}_1^*, \cdots, \tilde{h}_T^*)^{\top}\) can be obtained conventionally by minimizing the terminal AMSE and AMISE, respectively, i.e.,

$$\begin{align*}
\tilde{h}_t^* &= \arg\min_{H_t \in [0, \infty)} E_T(x; H_T), \\
\tilde{H}_T^* &= \arg\min_{H_T \in [0, \infty)^T} \int \mathcal{J}_x E_T(x; H_T) \tilde{W}(x) dx.
\end{align*}$$

A straightforward calculation leads to

$$\begin{align*}
\tilde{h}_t^* &= (C_h(x))^{1/5} N_T^{-1/5}, \\
\tilde{H}_T^* &= (C_h(x))^{1/5} N_T^{-1/5}
\end{align*}$$

(53)

for \(t = 1, \cdots, T\), where

$$C_h(x) = \frac{\Sigma_{m,x}}{4 (B_{m,x})^2}, \quad C_h = \frac{\int_{\mathcal{J}_x} \Sigma_{m,x} \tilde{W}(x) dx}{\int_{\mathcal{J}_x} \frac{1}{4} (B_{m,x})^2 \tilde{W}(x) dx}.$$  

The following remark shows the standard statistical convergence rate of \(\tilde{T}_T(x; J)\) under the optimal bandwidths.

#### Remark 4.2

Recall the error expressions in (48) and (51). We conclude that when the bandwidths are the optimal ones given in (53), the errors of \(\delta \tilde{T}_T \tilde{Y}_{x_T}^\delta\) and \(\tilde{T}_T(x; J)\) both enjoy the convergence rate of

$$O_p \left( N_T^{-2/5} \right) + O(|\Delta (G_x)|^{l+1}),$$

(54)

where the statistical error \(O_p(N_T^{-2/5})\) is of the optimal convergence one for nonparametric estimation; the numerical error \(O(|\Delta (G_x)|^{l+1})\) can be negligible compared with \(O_p(N_T^{-2/5})\) if the nodes in \(G_x\) are dense enough. More specifically, we have \(|\Delta (G_x)|^{l+1} = O(N_T^{-2/5})\) for uniformly spaced nodes satisfying \(#G_x \geq \delta_0 N_T^{2/5(l+5)}\), where \(\delta_0\) is a constant independent of \(N_T\). Such a condition on \(G_x\) is quite mild, because the degree \(l\) of LPI can be selected artificially, and the number \#\(G_x\) is only constrained by computing resources. We also remark that the convergence rate in (54) is obtained on the streaming data sets without any restrictions on the chunk size or chunk number.

#### 4.2.2 Practical sub-optimal bandwidths

It is usually closer to the real condition that the streaming data are endless or the terminal cumulative number \(N_T\) is unpredictable. In such a case, the theoretical optimal bandwidths in (53) are impractical. Thus, we have to find the sub-optimal bandwidths that do not rely on the information of future streaming data sets. To this end, we introduce the following lemma, which reveals the structure of the error function given in (52).

#### Lemma 4.5

The error function \(E_t\) defined in (52) has the decomposition:

$$E_t(x; H_t) = \left( \frac{N_t-1}{N_t} \right)^2 E_{t-1}(x; H_{t-1}) + \left( \frac{n_t}{N_t} \right)^2 E_t(x, h_t^\delta; H_{t-1})$$

with \(E_t(x, h_t^\delta; H_{t-1})\) defined by

$$E_t(x, h_t^\delta; H_{t-1}) = h_t^\delta (B_{m,x})^2 + \frac{1}{n_t h_t} \Sigma_{m,x} + 2 h_t^\delta \sum_{s=1}^{t} \frac{n_s h_t^2}{n_t} (B_{m,x})^2.$$

From Lemma 4.5, we can see that at \(t\)-th updating, the AMSE \(E_t\) can be expressed as a weighted sum of two parts: the first one \(E_{t-1}\) is the error of the last updating step, which has a fixed value at the current updating procedure; the second part \(E_t(x, h_t^\delta; H_{t-1})\) relies on \(h_t^\delta\) and \(H_{t-1}\). Note that \(H_{t-1}\) has been given in the previous updating procedure. It shows that only \(h_t^\delta\) need to be determined. Thus the optimal value of \(h_t^\delta\) should minimize \(E_t(x, h_t^\delta; H_{t-1})\) after the value of \(H_{t-1}\) is given. We thus define the sub-optimal variable and constant bandwidths respectively by

$$h_t^\delta = \arg\min_{h_t \in [0, \infty)} E_t(x, h_t; H_{t-1}^\delta)$$

$$h_t^\delta = \arg\min_{h_t \in [0, \infty)} \int_{\mathcal{J}_x} E_t(x, h_t; H_{t-1}^\delta) \tilde{W}(x) dx,$$

where \(H_{t-1}^\delta = (h_{t-1}^\delta(x), \cdots, h_{t-1}^\delta(x))^T\) and \(H_{t-1}^\delta = (h_{t-1}^\delta(x), \cdots, h_{t-1}^\delta(x))^T\) are the sub-optimal bandwidths selected for the first \(t-1\) data chunks. By straightforward calculation,
the sub-optimal bandwidths can be obtained by solving the following equations:

\[
\hat{h}_t^*(x) = \left\{ C_h(x) \right\}^{1/3} \left( \sum_{s=1}^{t-1} n_s \left( \hat{h}_s^*(x) \right)^2 + n_t \left( \hat{h}_t^*(x) \right)^2 \right)^{-1/3},
\]

\[
\hat{h}_t^* = C_h^{1/3} \left( \sum_{s=1}^{t-1} n_s \left( h_s^* \right)^2 + n_t \left( h_t^* \right)^2 \right)^{-1/3}.
\]

The sub-optimal bandwidths in (55) and (56) are actually given by two non-linear equations, which are not convenient for practical applications, especially in the scenario of the fast online-updating. To address this issue, we use \( h_{s-1}^* (x) \) and \( h_{t-1}^* \) to approximate the unknown \( h_s^* (x) \) and \( h_t^* \) on the right side of (55) and (56), respectively, which lead to the renewable bandwidths as

\[
\hat{h}_t (x) = \left\{ C_h (x) \right\}^{1/3} \left\{ S_{h,t} (x) \right\}^{-1/3},
\]

\[
\hat{h}_t = C_h^{1/3} \overline{S}_{h,t}^{-1/3}
\]

for \( t = 2, \ldots, T \), where

\[
S_{h,t} (x) = S_{h,t-1} (x) + n_t \left( \hat{h}_{t-1}^* (x) \right)^2,
\]

\[
S_{h,t} = S_{h,t-1} + n_t \left( \hat{h}_{t-1}^* \right)^2.
\]

with the initial values given by \( S_{h,1} (x) = S_{h,1} = 0, \hat{h}_1^* (x) = (C_h (x))^{1/5} n_1^{-1/5} \) and \( \hat{h}_t^* = (C_h (x))^{1/5} n_1^{-1/5} \).

The optimal and sub-optimal bandwidths in (53) and (57) all depend on unknown parameters. In practical applications, the unknown parameters can be estimated by cross validations on a validation data set consisting of samples collected from the first few data chunks. More details will be discussed in Sect. 5.

### 4.3 Optimal weight functions

Although the renewable WCQR estimators introduced in Sect. 3.3 can estimate \( m (x) \) and \( \sigma (x) \) with a convergence rate comparable to the oracle estimators, their weight functions are generally not optimal in terms of estimation variance. Based on the asymptotic distribution in Theorem 4.2, we can go a step further to find the optimal weight function in the sense of the associated renewable estimator \( m (x) \) or \( \sigma (x) \) having a minimized asymptotic variance. To this end, we introduce the set \( \mathcal{J}_m (\tau, \tau) \) (resp. \( \mathcal{J}_\sigma (\tau, \tau) \)) consisting of all the weight functions \( J (\cdot) : [0, 1] \rightarrow \mathbb{R} \) satisfying the constrains \( (C_{m1}, C_{m2}) \) (resp. \( (C_{\sigma1}, C_{\sigma2}) \)) and being square integrable with a support in \( [\tau, \tau] \subset (0, 1) \). Here \( \tau \) and \( \tau \) are pre-given parameters introduced in Assumption 4.4. By Theorem 4.2 and Corollary 4.3, the variance of \( \tilde{r} (x; J) \) can be expressed as

\[
\left( \frac{1}{N_t} \sum_{i=1}^{t} \frac{n_s}{\sum_{i=1}^{t} n_s} \right)^{-1} \text{Var} \{ \tilde{m} (x; J) \}
\]

\[
= \frac{k_0 \sigma^2 (x)}{f_x (x)} V (J) + O \left( |\Delta (G_x)|^{t+1} \right) + o (1)
\]

with \( V (\cdot) \) a functional defined by

\[
V (J) = \int_{[\tau, \tau]} \int_{[\tau, \tau]} S_f (\tau_1, \tau_2) d\tau_1 d\tau_2.
\]

For \( z = m (\cdot) \) or \( \sigma (\cdot) \), the optimal weight function \( J^*_z (\cdot) \) for estimating \( z (x) \) is given by the following functional minimization problem:

\[
J^*_z (\cdot) = \arg \min_{J \in \mathcal{J}_z (\tau, \tau)} V (J) \quad \text{for} \quad z = m, \sigma.
\]

By tools of variational analysis, we can obtain the closed-form expression of \( J^*_z (\cdot) \) given in the following theorem and corollary.

**Theorem 4.6** If \( f_x (\cdot) \) is twice differentiable, then the optimal weight function in (59) can be expressed by

\[
J^*_z (\tau) = I (\tau \leq \tau) \left( C_1 \Psi_1 (\tau) + C_2 \Psi_2 (\tau) \right),
\]

where \( \Psi_i (\cdot), i = 1, 2, \) are two basis function given by

\[
\Psi_i (\tau) = - \left( \delta_{ii} + \delta_{2i} I_4 \right) \log f_x^{i''} \left( F_x^{-1} (\tau) \right)
\]

with \( I_4 \) the identity function, i.e., \( I_4 (y) = y \) for \( y \in \mathbb{R} \); the coefficients \( C_1 \) and \( C_2 \) are selected to satisfy the corresponding conditions in (20) or (21) i.e.,

if \( z = m \), \( C_i = \delta_{ii} A_{12} - \delta_{2i} A_{11} \)

if \( z = \sigma \), \( C_i = \delta_{2i} A_{11} - \delta_{1i} A_{21} \)

with

\[
A_{kj} = \int_{[\tau, \tau]} \left( F_x^{-1} (\tau) \right)^k \Psi_j (\tau) d\tau
\]

for \( k = 0, 1, \) and \( j = 1, 2. \)

The following remark shows that Theorem 4.6 is an extension of existing results about the optimal weight functions in L-estimation.
Remark 4.3 Theorem 4.6 allows us to find the optimal weight function under the constraint $\text{Supp } (J(\cdot)) \subseteq [\tau, \bar{\tau}]$. This constraint is artificially introduced to control the error caused from LPI (see Remark 3.1). If we consider a special case of Theorem 4.6 where $[\tau, \bar{\tau}] = [0, 1]$, i.e., the constraint on $\text{Supp } (J(\cdot))$ is removed, then the optimal weight function $J^*_m(\cdot)$ (resp. $J^*_\sigma(\cdot)$) degenerates into $\Psi_1(\cdot)$ (resp. $\Psi_2(\cdot)$), which is just the optimal score functions for $m(x)$ (resp. $\sigma(x)$) shown in Portnoy and Koenker (1989), Koenker (2005).

The optimal weight function in (60) is deeply related to the distribution function and the PDF of the error, which is difficult to estimate especially in the case of streaming data. In the following corollary, we manage to express $J^*_\zeta(\cdot)$ using the CDF $F_{Y|X}(\cdot)$.

Corollary 4.7 Under Assumption 4.1, the optimal weight function in (60) can be expressed as

$$J^*_\zeta(\tau) = I(\bar{\tau} - \tau \leq \tau) (\tilde{C}_1 \tilde{\Psi}_1(\tau) + \tilde{C}_2 \tilde{\Psi}_2(\tau)), \quad (62)$$

where $\tilde{\Psi}_i(\cdot)$, $i = 1, 2$, are basis functions given by

$$\tilde{\Psi}_i(\tau) = -\int_{I_x} ( (\delta_{i1} + \delta_{i2})_i \log F_{Y|X}(\cdot) ) (F_{Y|X}^{-1}(\tau)) \, dx$$

and the coefficients $\tilde{C}_1$ and $\tilde{C}_2$ are selected to fulfill the corresponding conditions $(C_{\zeta1}, C_{\zeta2})$, i.e.,

$$\begin{align*}
\text{if } z &= m, \\
\tilde{C}_1 &= \frac{d_0 \tilde{A}_{12} - d_1 \tilde{A}_{02}}{\tilde{A}_{01} \tilde{A}_{12} - \tilde{A}_{02} \tilde{A}_{11}}, \\
\tilde{C}_2 &= \frac{-d_0 \tilde{A}_{11} + d_1 \tilde{A}_{01}}{\tilde{A}_{01} \tilde{A}_{12} - \tilde{A}_{02} \tilde{A}_{11}},
\end{align*}$$

$$\begin{align*}
\text{if } z &= \sigma, \\
\tilde{C}_1 &= \tilde{A}_{01} + \tilde{C}_2 \tilde{A}_{02} = 0, \\
\tilde{C}_2 &= \tilde{C}_1 \tilde{D}_{1,1} + 2 \tilde{C}_1 \tilde{C}_2 \tilde{D}_{1,2} + \tilde{C}_2^2 \tilde{D}_{2,2} = d_3^2,
\end{align*}$$

with $d_0 = 1$, $d_1 = E[ W(X) X ]$ and

$$d_3^2 = E\left[ W(X) (Y^2 - m^2(X)) \right],$$

$$\tilde{A}_{kj} = \int_{I_x} \int_{\mathbb{R}} ( y W(x) f_X(x) )^k \times \tilde{\Psi}_j (F_{Y|X}(y)) dF_{Y|X}(y) \, dx$$

for $k = 0, 1$ and $j = 1, 2$.

Corollary 4.7 suggests a renewable estimation method for the optimal weight function based on the interpolated empirical CDF given in (14). Specifically speaking, the estimator of $J^*_\zeta(\cdot)$ can be given by (62) with the unknown $F_{Y|X}(\cdot)$ and its derivatives replaced by $\int_{I_x} \tilde{F}_{Y|X,t}(\cdot)$ and its corresponding derivatives, and the involved constants $\tilde{C}_i$ can be estimated by renewable procedures similar to the ones introduced in (29) and (39).

5 Numerical experiments

In this section, we conduct simulation studies and real data analyses to verify performance of various estimators involved in this paper.

In the numerical experiments, all the involved kernel functions are taken as the Epanechnikov kernel, i.e., $K(z) = \max \{ 0, 3/4 (1 - z^2) \}$. To implement the renewable WCQR estimation, 3rd-degree LPI is used to obtain the interpolated empirical CDF. To obtain the LPI nodes, the set $G_s$ is formed by 100 grid points evenly distributed over the intervals $I_s$. And for each $x_i$ in $G_s$, the set $G_{x_i}$ consists of the points $y_{ij}$ given by

$$y_{ij} = \arg \min_{y \in \mathbb{R}} \sum_{q : x_q \in A_{x_i}(x, D_{val})} \rho_{\tau_j} (Y_q - y)$$

for $\tau_j = j/40$ and $j = 1, \ldots, 40$. Here $D_{val}$ is a validation data set containing the first 2000 samples collected from the data stream. And the function $N_{\tau_j}$ is defined in Sect. 2.2 with $k$ empirically selected as max $\{ 0.1 \#D_{val}, \#G_{x_i} \}$.

For implementations, we need to artificially determine a terminal time $T$ for the streaming data sets. Depending on whether or not the cumulative number $N_T$ is supposed to be known, we consider two bandwidth selectors: the “oracle” optimal constant bandwidth $h_0^*$ from (53) and the renewable constant bandwidth $\tilde{h}_r$ from (57). The unknown constant $C_h$ in (53) and (57) is replaced by its estimator $\tilde{C}_h$ obtained by 10-fold cross validation on the validation data set $D_{val}$.

We mainly discuss the performance of the following three proposed renewable WCQR estimators:

- $\tilde{r}_{atm}$: It is the renewable NTM $\tilde{r}_T(x; J_{m,0.5})$ given by (15) with the weight function $J_{m,0.5}$ given in (27) with $\alpha = 0.1$; the bandwidth is selected as the renewable $\tilde{h}_r$ given in (57) with $C_h$ replaced by $\tilde{C}_h$.
- $\tilde{r}_{bctm}$: It is the renewable BCTM $\tilde{r}_T(x; J_{m,\tilde{h}_r})$ given by (15) with the weight function $J_{m,\tilde{h}_r}$ given in (27) with $\alpha = 0.1$, $\tilde{h}_r$ obtained by (29) and $W(\cdot)$ given by (25); the bandwidth selector is the same with that of $\tilde{r}_{atm}$.
- $\tilde{r}_{rtsd}$: It is the renewable RTSD $\tilde{r}_T(x; J_{\sigma,\tilde{h}_r})$ given by (15) with the weight function $J_{\sigma,\tilde{h}_r}$ given in (37) with $\alpha = 0.1$, $\tilde{h}_r$ obtained by (39) and $W(\cdot)$ given by (25); the bandwidth selector is the same with that of $\tilde{r}_{atm}$.

For comparison, we introduce the following benchmark estimators:

- $\tilde{r}_{atm}^*$, $\tilde{r}_{bctm}^*$ and $\tilde{r}_{rtsd}^*$: They are the oracle counterparts of $\tilde{r}_{atm}$, $\tilde{r}_{bctm}$ and $\tilde{r}_{rtsd}$, respectively, i.e., all of them are computed on the full data $\cup_{t \leq T} D_t$ with the bandwidth selected as the estimated oracle optimal constant bandwidths, i.e., $h = \tilde{C}_h N_T^{-1/5}$.
• \( \hat{r}_{\text{ntm}}^{\text{w}} \), \( \hat{r}_{\text{bctm}}^{\text{w}} \) and \( \hat{r}_{\text{tsd}}^{\text{w}} \): They are the simple-average counterparts of \( \hat{r}_{\text{ntm}}, \hat{r}_{\text{bctm}} \) and \( \hat{r}_{\text{tsd}} \), respectively, which are obtained by simply averaging the corresponding local estimators computed on the data chunks \( D_1, \ldots, D_T \). For example, \( \hat{r}_{\text{ntm}}^{\text{w}}(x) = 1/T \sum_{t=1}^{T} \hat{r}_{\text{ntm}}^{t}(x) \), where \( \hat{r}_{\text{ntm}}^{t} \) is the analogue of \( \hat{r}_{\text{ntm}} \) computed on \( D_t \). Here all the bandwidths are selected as the estimated local optimal constant bandwidths, i.e., \( h_t = \hat{C}_h n_t^{-1/5} \).

• \( \hat{r}_{\text{kcqr}} \): It is the oracle CQR estimator proposed by Kai et al. (2010) and computed on the full data \( U_{\text{t}} \). Specifically, it is the estimator \( \hat{m}_{\text{kcqr}}(x) \) given by (2) with \( \tau_t = i/(q + 1) \), \( q = 9 \), \( \hat{Q}_{Y|x}(\tau_t) = \hat{a} \) and \( (\hat{a}, \hat{b}) \) the minimizer of

\[
\sum_{t=1}^{T} \sum_{j=1}^{n_t} \rho_{\text{c}}(Y_{ij} - a - b(X_{ij} - x)) K_h(x_{ij} - x)
\]

over \((a, b) \in \mathbb{R}^2 \). Here the bandwidth \( h \) is selected as the oracle optimal constant bandwidths, i.e., \( h = C'_h N_T^{-1/5} \) with the constant \( C'_h \) estimated by 10-fold cross validation on \( D_{\text{val}} \). In our experiments, \( \hat{r}_{\text{kcqr}} \) is a competitor of the proposed \( \hat{r}_{\text{ntm}} \), because both of them are CQR-type estimators for \( m(x) \); Moreover, both of them are consistent for symmetric models, and are inconsistent for asymmetric models; see, e.g., (Kai et al. 2010; Sun et al. 2013).

• \( \hat{r}_{\text{swcqr}}^{\text{w}} \): It is the oracle WCQR estimator proposed by Sun et al. (2013) and computed on the full data \( U_{\text{t}} \). Specifically, it is the estimator \( \hat{m}_{\text{swcqr}}(x) \) given by (3) with \( \tau_t = i/(q + 1) \), \( q = 9 \) and \( \hat{Q}_{Y|x}(\tau_t) \) obtained by the same ways as \( \hat{r}_{\text{kcqr}} \). The weights \( \omega_{ij} \) in (3) are selected as the optimal ones given by Sect. 3.2 in Sun et al. (2013). In our experiments, \( \hat{r}_{\text{swcqr}}^{\text{w}} \) is a competitor of the proposed \( \hat{r}_{\text{bctm}} \), because both of them are CQR-type estimators for \( m(x) \), and both of them are consistent estimators no matter whether the model (1) is symmetric or not.

• \( \hat{r}_{\text{nw}}^{\text{w}} \) and \( \hat{r}_{\text{nw}}^{\text{w}} \): They are the oracle Nadaraya-Watson (NW) estimators for \( m(x) \) and \( \sigma(x) \), respectively, which are computed on the full data \( U_{\text{t}} \), i.e.,

\[
\hat{r}_{\text{nw}}^{\text{w}}(x) = \frac{\sum_{t=1}^{T} \sum_{j=1}^{n_t} Y_{ij} K_h(x_{ij} - x)}{\sum_{t=1}^{T} \sum_{j=1}^{n_t} K_h(x_{ij} - x)}, \tag{63}
\]

and

\[
(\hat{r}_{\text{nw}}^{\text{w}}(x))^2 = \frac{\sum_{t=1}^{T} \sum_{j=1}^{n_t} (Y_{ij} - \hat{r}_{\text{nw}}^{\text{w}}(x))^2 K_h(x_{ij} - x)}{\sum_{t=1}^{T} \sum_{j=1}^{n_t} K_h(x_{ij} - x)}, \tag{64}
\]

where the bandwidth is selected as the oracle optimal constant bandwidths, i.e., \( h = C'_h N_T^{-1/5} \) with the constant \( C'_h \) estimated by 10-fold cross validation on \( D_{\text{val}} \). The above-mentioned \( \hat{r}_{\text{nw}}^{\text{w}} \) and \( \hat{r}_{\text{nw}}^{\text{w}} \) are classical estimators for \( m(\cdot) \) and \( \sigma(\cdot) \), respectively, and both of them enjoy consistency no matter whether the model (1) is symmetric or not. In our experiments, \( \hat{r}_{\text{nw}}^{\text{w}} \) and \( \hat{r}_{\text{nw}}^{\text{w}} \) are competitors of the proposed \( \hat{r}_{\text{bctm}} \) and \( \hat{r}_{\text{tsd}} \), respectively. Particularly, for symmetric models, \( \hat{r}_{\text{nw}}^{\text{w}} \) is also a competitor of the proposed \( \hat{r}_{\text{ntm}} \).

In the experiments, all the algorithms are implemented in Python 3.9.12 with the runtimes obtained by using the standard Python function: time.perf_counter.

### 5.1 Simulation studies

In the simulation studies, we will consider various experiment conditions, such as homoscedastic or heteroscedastic models, and symmetric or asymmetric errors. To this end, we consider the following two models:

Model 1: \( Y = \sin(2X) + 2 \exp \left(-16X^2\right) + 0.5e \)

with \( X \sim N(0, 1) \), \( I_s = [-1.5, 1.5] \).

Model 2: \( Y = \sin(2\pi X) + (2 + \cos(2\pi X)) \epsilon \)

with \( X \sim U(0, 1) \), \( I_s = [0, 1] \).

where Model 1 is homoscedastic and adopted from Fan and Gijbels (1992), and Model 2 is heteroscedastic and adopted from Kai et al. (2010) We consider various kinds of distributions of \( \epsilon \). Moreover, we also use the mixtures of two error distributions to model so-called contaminated data. Specifically, a mixture distribution is chosen as \( 0.95F_\epsilon + 0.05F_\lambda, \epsilon \) with a multiplying factor \( \lambda \), where \( F_\epsilon \) is the distribution function of \( \epsilon \). If without special statement, all the distributions of \( \epsilon \) involved in the simulations are normalized such that \( \mathbb{E}[\epsilon] = 0 \) and \( \text{Var}(\epsilon) = 1 \). Based on the combinations of different models and error distributions, we consider the following four examples:

• Example 1a: the Model 1 with various symmetric error distributions;
• Example 1b: the Model 1 with various asymmetric error distributions;
• Example 2a: the Model 2 with various symmetric error distributions;
• Example 2b: the Model 2 with various asymmetric error distributions.

To model the streaming data, we generate the full data of size \( N_T \) from the considered models and equally divide the full data into \( T \) data chunks.

For each estimation, the number of replications in the simulation is designed as 500. The performance of any estimator.
\( \hat{g}( \cdot ) \) of a function \( g( \cdot ) \) is evaluated by the average squared errors (ASEs) defined by

\[
\text{ASE}( \hat{g} ) = \frac{1}{\#G_s} \sum_{x_i \in G_s} | \hat{g}( x_i ) - g( x_i ) |^2.
\]

To compare the performance of two estimators \( \hat{g}_1 \) and \( \hat{g}_2 \), we use the ratio of average squared errors (RASEs):

\[
\text{RASE}( \hat{g}_1, \hat{g}_2 ) = \frac{\text{ASE}( \hat{g}_1 )}{\text{ASE}( \hat{g}_2 )}.
\]

### 5.1.1 Scenario 1: streaming data with varying chunk size

We first discuss the influence of data partitioning on our WCQR estimators. To this end, we fix the full sample size \( NT = 10^5 \) and successively take the chunk number \( T = 10, 10^2, 10^3, 10^4 \), resulting the chunk sizes \( n_t = 10^4, 10^3, 10^2, 10 \), respectively. The multiplying factor is \( \lambda = 1 \), i.e., the streaming data is not contaminated. Then we test the WCQR estimators in Example 1a and 1b with the associated simple-average estimators and oracle estimators used as benchmarks. The relevant RASEs and computational times are reported in Table 1 and Table 2, respectively. As stated before, \( \hat{r}_{\text{ntm}} \) and \( \hat{r}_{\text{kcqr}} \) are generally inconsistent for asymmetric models. Thus the indicator RASE(\( \hat{r}_{\text{kcqr}}, \hat{r}_{\text{ntm}} \)) is not applicable for Example 1b, and we do not report it in Table 1.

For the results in Table 1, we have the following discussions:

(i) From RASE(\( \hat{r}_{\text{ntm}}, \hat{r}_{\text{ntm}} \)), RASE(\( \hat{r}_{\text{bctm}}, \hat{r}_{\text{bctm}} \)) and RASE(\( \hat{r}_{\text{rstd}}, \hat{r}_{\text{rstd}} \)) in Examples 1a and 1b, we can see that the renewable WCQR estimators \( \hat{r}_{\text{ntm}}, \hat{r}_{\text{bctm}} \) and \( \hat{r}_{\text{rstd}} \) perform well with all the associated RASEs not only insensitive to chunk size \( n_t \) but also closed to 1. On the contrary, the simple-average estimators \( \hat{r}_{\text{ntm}}, \hat{r}_{\text{bctm}} \) and \( \hat{r}_{\text{rstd}} \) are susceptible to the chunk size. For small chunk sizes, i.e., \( n_t < 10^3 \), their performances are significantly inferior than that of the oracle estimators. In the extreme case \( n_t = 10 \), the renewable WCQR estimators still works well, but the simple-average estimators can not give results because the data chunk is too small to compute the local estimators.

(ii) From RASE(\( \hat{r}_{\text{kcqr}}, \hat{r}_{\text{ntm}} \)) and RASE(\( \hat{r}_{\text{nw}}, \hat{r}_{\text{ntm}} \)) in Example 1a, we can see that for symmetric models, the renewable NTM \( \hat{r}_{\text{ntm}} \) performs as well as the oracle estimators \( \hat{r}_{\text{kcqr}} \) and \( \hat{r}_{\text{nw}} \) despite the values of the chunk size \( n_t \). Meanwhile, the RASE(\( \hat{r}_{\text{nw}}, \hat{r}_{\text{ntm}} \)) in Example 1b shows that \( \hat{r}_{\text{ntm}} \) is not applicable for asymmetric models.

(iii) By RASE(\( \hat{r}_{\text{swcqr}}, \hat{r}_{\text{bctm}} \)) and RASE(\( \hat{r}_{\text{rstd}}, \hat{r}_{\text{rstd}} \)) in Examples 1a and 1b, we can see that for all involved chunk sizes, the renewable BCTM \( \hat{r}_{\text{bctm}} \) and RTSD \( \hat{r}_{\text{rstd}} \) both perform well compared with their oracle competitors.

Moreover, when the error distribution is nonnormal, i.e., Pareto(3), the renewable \( \hat{r}_{\text{rstd}} \) even shows significant superiority over the oracle \( \hat{r}_{\text{rstd}} \).

By the results in Table 2, we can find that the runtimes for the proposed renewable estimators \( \hat{r}_{\text{ntm}}, \hat{r}_{\text{bctm}} \) and \( \hat{r}_{\text{rstd}} \) all depend on the chunk size \( n_t \) or the chunk number \( T = NT/n_t \). Even in the worst case, i.e., \( n_t \equiv 10 \), the above-mentioned renewable estimators are more computational efficient than the oracle CQR-typed estimators \( \hat{r}_{\text{kcqr}} \) and \( \hat{r}_{\text{swcqr}} \). This is because our renewable estimators have closed-form expressions without the need of solving any optimization problems. But \( \hat{r}_{\text{kcqr}} \) and \( \hat{r}_{\text{swcqr}} \) both rely on optimization algorithms to obtain the quantile estimators \( \hat{Q}_{y|x}(\tau_i) \) in (2) and (3).

We also notice that the oracle NW estimators \( \hat{r}_{\text{nw}} \) and \( \hat{r}_{\text{nw}} \) are the most computationally efficient, because their closed-form expressions (63) and (64) are quite simple.

By the above discussions, we conclude that our renewable algorithm is computationally efficient and enjoys desirable performance robust to the chunk size of the streaming data. Our renewable WCQR estimators are comparable to their oracle competitors.

### 5.1.2 Scenario 2: models with contaminated streaming data

We turn to test the robustness and model adaptiveness of our WCQR estimators. We successively take various multiplying factors \( \lambda = 1, 3, 5, 10 \) for contaminated streaming data. Since all the involved estimators are impervious to the data partitioning, we only consider a fixed chunk size \( n_t = 100 \) with the full data size \( NT = 10^5 \). Then we test the estimators in the four examples and report the relevant RASEs in Table 3.

For the results in Table 3, we have the following discussions:

(i) In Examples 1a and 2a, the model is symmetric, and by the values of RASE(\( \hat{r}_{\text{nw}}, \hat{r}_{\text{ntm}} \)) and RASE(\( \hat{r}_{\text{nw}}, \hat{r}_{\text{bctm}} \)) the two estimators \( \hat{r}_{\text{ntm}} \) and \( \hat{r}_{\text{bctm}} \) show similar behaviors. Specifically, when \( \gamma = 1 \), i.e., the error is normal without contaminations, both of them are slightly inferior than the oracle NW estimator \( \hat{r}_{\text{nw}} \). However, when \( \lambda > 1 \), i.e., the streaming data are contaminated, both of them perform \( \hat{r}_{\text{nw}} \). Moreover, the values of RASE(\( \hat{r}_{\text{nw}}, \hat{r}_{\text{ntm}} \)) and RASE(\( \hat{r}_{\text{nw}}, \hat{r}_{\text{bctm}} \)) increase as \( \lambda \) is increasing, which suggests that compared with the NW estimator, the NTM and BCTM are more robust to data contaminations. We also notice that even if \( \lambda \) is large, there is no obvious gap between the performance of \( \hat{r}_{\text{ntm}} \) and \( \hat{r}_{\text{bctm}} \). This justifies our claim in Remark 3.6 that the robustness of the BCTM is not susceptible to the non-robust estimation of \( E_{\text{WY}} \).
We focus on the values of RASE $\hat{r}_{\text{ntm}}$ and RASE $\hat{r}_{\text{bctm}}$ in Examples 1b and 2b, where the model is asymmetric. Contrary to the case of symmetric models, the behaviors between the NTM and the BCTM are quite different. All the values of RASE $\hat{r}_{\text{ntm}}$, $\hat{r}_{\text{bctm}}$ are closed to zero, indicating that the NTM is far inferior to the NW estimator and it can be inconsistent for asymmetric models. While, the values of RASE $\hat{r}_{\text{ntm}}$, $\hat{r}_{\text{bctm}}$ suggest that the BCTM still works well and even outperforms the NW estimators when the data is contaminated.

The above results show that under the weight selection criteria in Sect. 3.2, our renewable WCQR estimator is adaptive to symmetric and asymmetric models.

By the values of RASE $\hat{r}_{\text{ntm}}$, $\hat{r}_{\text{bctm}}$ in Example 1a and 2a, for all presented values of $\lambda$, the NTM $\hat{r}_{\text{ntm}}$ achieves a performance close to the oracle estimator $\hat{r}_{\text{cqr}}$, which means that the robustness of the NTM is comparable to its CQR-typed competitor $\hat{r}_{\text{cqr}}$. From the values of RASE $\hat{r}_{\text{ntm}}$, $\hat{r}_{\text{bctm}}$ in the four examples, we can find that the BCTM $\hat{r}_{\text{bctm}}$ behave slightly better than $\hat{r}_{\text{swcqr}}$.

(iii) We focus on the values of RASE $\hat{r}_{\text{ntm}}$ and RASE $\hat{r}_{\text{bctm}}$ in Examples 1a and 2a, where the model is asymmetric. Contrary to the case of symmetric models, the behaviors between the NTM and the BCTM are quite different. All the values of RASE $\hat{r}_{\text{ntm}}$, $\hat{r}_{\text{bctm}}$ are closed to zero, indicating that the NTM is far inferior to the NW estimator and it can be inconsistent for asymmetric models. While, the values of RASE $\hat{r}_{\text{ntm}}$, $\hat{r}_{\text{bctm}}$ suggest that the BCTM still works well and even outperforms the NW estimators when the data is contaminated.

The above results show that under the weight selection criteria in Sect. 3.2, our renewable WCQR estimator is adaptive to symmetric and asymmetric models.

(iv) By the values of RASE $\hat{r}_{\text{ntm}}$, $\hat{r}_{\text{bctm}}$ in Example 1a and 2a, for all presented values of $\lambda$, the NTM $\hat{r}_{\text{ntm}}$ achieves a performance close to the oracle estimator $\hat{r}_{\text{cqr}}$, which means that the robustness of the NTM is comparable to its CQR-typed competitor $\hat{r}_{\text{cqr}}$. From the values of RASE $\hat{r}_{\text{ntm}}$, $\hat{r}_{\text{bctm}}$ in the four examples, we can find that the BCTM $\hat{r}_{\text{bctm}}$ behave slightly better than $\hat{r}_{\text{swcqr}}$.

Springer
Now we discuss the values of $RASE$ for $\lambda > 3$. The reason may be that the weight selection of $\hat{r}_{swcqr}$ relies on local least square estimations (see Sect. 3.2 in [Sun et al. 2013]), which decreases the robustness of $\hat{r}_{swcqr}$.

(v) Now we discuss the values of $RASE(\hat{r}_{nw}, \hat{r}_{nt})$ in the four examples. As $\lambda$ is increasing, the RASEs between $\hat{r}_{nw}$ and $\hat{r}_{nt}$ show an increasing trend with the values uniformly larger than 1 when $\lambda > 3$. This means that our estimator $\hat{r}_{nt}$ is more robust than $\hat{r}_{nw}$ and enjoy more advantages when the streaming data are contaminated.

5.1.3 Scenario 3: models with nonnormal error distributions

We focus on the performance of our WCQR estimators for nonnormal error distributions. We still consider a fixed chunk size $n_t = 100$ with the full data size $N_T = 10^5$. The multiplying factor is $\lambda = 1$, i.e., there is no contaminated data. We test the WCQR estimators and their competitors in the four examples with nonnormal error distributions. The relevant RASEs are reported in Table 4.

From Table 4, we have the following finds:

(i) We focus on the values of $RASE(\hat{r}_{nw}, \hat{r}_{nt})$, $RASE(\hat{r}_{nw}, \hat{r}_{bctm})$ and $RASE(\hat{r}_{nw}, \hat{r}_{nts})$ in the four examples. For symmetric models with nonnormal errors, i.e., Examples

Table 3 The performance comparison between the renewable WCQR estimators and the oracle benchmark estimators under various models with contaminated streaming data; the full sample size is $N_T = 10^5$ and the chunk size is $n_t = 1000$

| Example | Error distribution | $\lambda$ | $RASE(\hat{r}_{nw}, \hat{r}_{nt})$ | $RASE(\hat{r}_{nw}, \hat{r}_{bctm})$ | $RASE(\hat{r}_{nw}, \hat{r}_{nts})$ | $RASE(\hat{r}_{nw}, \hat{r}_{nt})$ | $RASE(\hat{r}_{nw}, \hat{r}_{bctm})$ | $RASE(\hat{r}_{nw}, \hat{r}_{nts})$ |
|---------|-------------------|-----------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| 1a      | N(0, 1)           | 1         | 0.9185 0.2585                    | 0.9838 0.0336                    | 0.9149 0.2587                    | 0.9819 0.0347                    | 0.9354 0.0587                    |
|         |                   | 3         | 1.0230 0.2458                    | 1.1740 0.0884                    | 1.0007 0.2462                    | 1.1702 0.0859                    | 0.9903 0.0335                    |
|         |                   | 5         | 0.9783 0.2633                    | 1.5667 0.2285                    | 1.1777 0.2707                    | 1.5492 0.2247                    | 1.0079 0.0175                    |
|         |                   | 10        | 0.9441 0.2226                    | 3.0526 0.6052                    | 1.2702 0.3285                    | 2.9787 0.5955                    | 1.0227 0.0123                    |
| 1b      | F(10, 6)          | 1         | – –                             | 0.0774 0.0191                    | 0.8109 0.1646                    | 1.3746 0.2500                    | 0.9998 0.1701                    |
|         |                   | 3         | 0.0853 0.0197                    | 1.0338 0.1558                    | 1.6597 0.3949                    | 1.3974 0.3543                    |
|         |                   | 5         | 0.1831 0.1020                    | 1.0518 0.2295                    | 2.9289 0.3439                    | 2.6362 0.8594                    |
|         |                   | 10        | 0.2862 0.0763                    | 1.1557 0.2111                    | 4.2620 1.3305                    | 3.1271 0.9163                    |
| 2a      | N(0, 1)           | 1         | 0.9275 0.2329                    | 0.9943 0.0266                    | 0.9542 0.2842                    | 0.9877 0.0779                    | 0.8597 0.0863                    |
|         |                   | 3         | 0.9583 0.2062                    | 1.1058 0.1281                    | 0.9555 0.2299                    | 1.0970 0.1490                    | 1.0141 0.0198                    |
|         |                   | 5         | 0.9220 0.1853                    | 1.3614 0.1894                    | 1.1166 0.2143                    | 1.4573 0.2021                    | 1.2385 0.0111                    |
|         |                   | 10        | 0.8269 0.1467                    | 2.8376 0.6976                    | 1.2839 0.2484                    | 2.2326 0.4564                    | 1.3569 0.0560                    |
| 2b      | F(10, 6)          | 1         | – –                             | 0.1536 0.0232                    | 0.8512 0.1251                    | 1.2145 0.1340                    | 0.8954 0.1013                    |
|         |                   | 3         | 0.0949 0.0260                    | 0.9415 0.1476                    | 1.3247 0.1420                    | 1.0559 0.1851                    |
|         |                   | 5         | 0.1363 0.0367                    | 1.0942 0.1724                    | 1.9310 0.3144                    | 1.4738 0.5007                    |
|         |                   | 10        | – –                             | 0.4251 0.1283                    | 1.1770 0.1900                    | 2.4008 0.4228                    | 2.1477 1.0751                    |

Table 4 The performance comparison between the renewable WCQR estimators and the oracle benchmark estimators under various models and error distributions, the full sample size is $N_T = 10^5$, the chunk size is $n_t = 1000$ and the multiplying factor of contaminated data is $\lambda = 1$

| Example | Error distribution | RASE($\hat{r}_{nw}$, $\hat{r}_{nt}$) | RASE($\hat{r}_{nw}$, $\hat{r}_{nt}$) | RASE($\hat{r}_{nw}$, $\hat{r}_{bctm}$) | RASE($\hat{r}_{nw}$, $\hat{r}_{nts}$) | RASE($\hat{r}_{nw}$, $\hat{r}_{nt}$) | RASE($\hat{r}_{nw}$, $\hat{r}_{bctm}$) | RASE($\hat{r}_{nw}$, $\hat{r}_{nts}$) |
|---------|-------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| 1a      | Standard Laplace  | 1.0037 0.2496                    | 1.2386 0.1168                    | 0.9285 0.1840                    | 1.2078 0.1138                    | 1.4035 0.2036                    |
|         | t(3)              | 0.9272 0.2299                    | 1.4613 0.1751                    | 0.9165 0.2655                    | 1.3697 0.1455                    | 2.4456 1.1808                    |
| 1b      | F(4, 6)           | – –                             | 0.0666 0.0300                    | 0.9451 0.2533                    | 1.2383 0.2954                    | 1.7548 1.1452                    |
|         | Pareto(3)         | – –                             | 0.0585 0.0133                    | 1.1875 0.3217                    | 1.2152 0.2707                    | 3.7343 1.7048                    |
| 2a      | Standard Laplace  | 0.9096 0.1319                    | 1.1751 0.1351                    | 0.8898 0.2700                    | 1.2158 0.1373                    | 0.9295 0.1784                    |
|         | t(3)              | 0.8935 0.2539                    | 1.3172 0.1785                    | 0.9146 0.2919                    | 1.3620 0.1710                    | 2.7581 1.1581                    |
| 2b      | F(10, 6)          | – –                             | 0.1536 0.0232                    | 0.8512 0.1251                    | 1.2145 0.1340                    | 0.9954 0.1013                    |
|         | Lognorm(0, 1)     | – –                             | 0.0705 0.0151                    | 0.9215 0.1596                    | 1.1176 0.1286                    | 2.4585 1.1983                    |
Table 5  The relations studied in Real Data Example

| Example | X    | Y    | #\(D_{\text{train}}\) | #\(D_{\text{test}}\) |
|---------|------|------|------------------------|----------------------|
| 3a      | DEWP | O3   | 305572                 | 101808               |
| 3b      | WSPM | PM10 | 311146                 | 103146               |

1 and 2a, the CQR-typed estimators \(\hat{\gamma}_{\text{ntm}}\), \(\hat{\gamma}_{\text{bctm}}\) and \(\hat{\gamma}_{\text{nw}}\) can be more efficient than the NW-typed estimators \(\hat{\gamma}_{\text{nw}}^*\) and \(\hat{\gamma}_{\text{nw}}^\text{nwsd}\). The phenomenon is in line with the feature of CQR method in (Zou and Yuan 2008; Kai et al. 2010). For asymmetric models, i.e., Examples 1b and 2b, benefitting from the model adaptiveness mentioned in the last scenario, \(\hat{\gamma}_{\text{bctm}}\) and \(\hat{\gamma}_{\text{nw}}\) maintain their advantage over \(\hat{\gamma}_{\text{nw}}^*\) and \(\hat{\gamma}_{\text{nw}}^\text{nwsd}\). And as expected, the NTM \(\hat{\gamma}_{\text{ntm}}\) does not work because of the nonnegligible bias arising in asymmetric models.

(iii) By the values of RASE(\(\tilde{\gamma}_{\text{kcqr}}\), \(\tilde{\gamma}_{\text{ntm}}\)) and RASE(\(\tilde{\gamma}_{\text{swcqr}}\), \(\tilde{\gamma}_{\text{bctm}}\)) in the four examples, we conclude that the NTM and the BCTM are also comparable to their oracle CQR-typed competitors for all presented nonnormal error distributions.

5.2 Real data example

For case study, we apply our method to the Beijing Multi-Site Air-Quality Data set from the UCI machine learning repository. This data set consists of hourly data about 6 main air pollutants and 6 relevant meteorological variables collected from 12 nationally-controlled air-quality monitoring sites in Beijing, China. The observational data cover the time period from March 1st, 2013 to February 28th, 2017, and the 420,768 observed values of each variable. Our goal is to fit the relationship between the main air pollutants and the relevant meteorological variables in the dataset.

From the data set, we select two pairs of air pollutants and meteorological variables as the response variable \(Y\) and the covariate \(X\) in model (1), and then we obtain two examples listed in Table 5. In Table 5, the terms DEWP, O3, WSPM and PM10 are abbreviations to dew point temperature (degree Celsius), O3 concentration (ug/m\(^3\)), wind speed (m/s) and PM10 concentration (ug/m\(^3\)), respectively.

To test the performance of the involved estimators, we drop the data that suffer from data missing and then divide the remainder data set into training set \(D_{\text{train}}\) and testing set \(D_{\text{test}}\). The set \(D_{\text{train}}\) consists of the data collected before March 1st, 2017, which are used to obtain the involved estimators. And the set \(D_{\text{test}}\) consists of the data collected after March 1st, 2017, which are used to test the involved estimators. The number of samples in \(D_{\text{train}}\) and \(D_{\text{test}}\) are listed in Table 5.

To model the data stream, the data in \(D_{\text{train}}\) are revealed to the algorithm chronologically by chunks. According to the reality, we consider three different sizes of data chunks, where the data chunks are respectively formed by monthly, daily and hourly data in \(D_{\text{train}}\). To simulate the data contamination, we randomly select 5% samples \((X_{ti}, Y_{ti})\) from \(D_{\text{train}}\) and replace by \((X_{ti}, Y_{ti} + \eta_{ti})\), where \(\eta_{ti}\) are random numbers sampled from \(N(0, \gamma^2\sigma^2)\) with \(\sigma\) the sample standard deviation of all \(Y_{ti}\) in \(D_{\text{train}}\), and \(\gamma\) a scale factor controlling the intensity of data contaminations.

For an estimator \(\hat{g}(\cdot)\), its prediction accuracy is described by the root mean square error (RMSE) and the mean absolute error (MAE) on \(D_{\text{test}}\), namely

\[
\text{RMSE}(\hat{g}) = \sqrt{\frac{1}{n_{\text{test}}} \sum_{(X_i, Y_i) \in D_{\text{test}}} (Y_i - \hat{g}(X_i))^2},
\]

\[
\text{MAE}(\hat{g}) = \frac{1}{n_{\text{test}}} \sum_{(X_i, Y_i) \in D_{\text{test}}} |Y_i - \hat{g}(X_i)|,
\]

where \(n_{\text{test}}\) is the number of observations in \(D_{\text{test}}\).

The RMSEs and MAEs of the involved estimators are reported in Tables 6 and 7, respectively. From Tables 6 and 7, we have the following findings:

(i) We focus on the RMSEs and MAEs under different data chunk levels. In most cases, the renewable estimators \(\hat{\gamma}_{\text{ntm}}\) and \(\hat{\gamma}_{\text{bctm}}\) show performance impervious to the data partitioning. Specifically, their RMSEs and MAEs are insensitive to the data chunk levels and are almost the same with that of the oracle counterparts. Moreover, their errors are also close to their oracle CQR-typed competitors \(\hat{\gamma}_{\text{kcqr}}\) and \(\hat{\gamma}_{\text{swcqr}}\). On the contrary, the simple-average estimators \(\hat{\gamma}_{\text{bctm}}\) is susceptible to the data chunk levels, and their performance deteriorates significantly when the data chunks becomes smaller. This shows that our renewable WCQR estimation can overcome the challenge arising from the data partitioning, and our renewable estimators enjoy asymptotic properties comparable to that of the oracle estimator obtained on the full data set.

(ii) As the scale factor \(\gamma\) increasing, the errors of the renewable NTM \(\hat{\gamma}_{\text{ntm}}\) are insensitive to \(\gamma\) and always close to the error of its oracle CQR-typed competitor \(\hat{\gamma}_{\text{kcqr}}\). For most cases with \(\gamma > 0\), i.e., the data is contaminated, \(\hat{\gamma}_{\text{ntm}}\) is superior than \(\hat{\gamma}_{\text{nw}}\) in both terms of RMSE and MAE. Moreover, this advantage is enlarged when \(\gamma\) is relatively large. In general, the above results suggest that our renewable WCQR estimation can achieve robustness for contaminated streaming data.

(iii) As the scale factor \(\gamma\) increasing, the behavior of the renewable BCTM \(\hat{\gamma}_{\text{bctm}}\) is similar to its CQR-typed competitor \(\hat{\gamma}_{\text{swcqr}}\). When \(\gamma = 0\), i.e., there is no contaminated

1 https://archive-beta.ics.uci.edu/ml/datasets/beijing+multi+site+air+quality+data
### Table 6 The RMSEs of various estimators in fitting the test set from the Beijing Multi-Site Air-Quality Data set

| Example | $\gamma$ | Data chunk levels | $\hat{\gamma}_{kcqr}$ | $\hat{\gamma}_{ntm}$ | $\hat{\gamma}_{swcqr}$ | $\hat{\gamma}_{bctm}$ | $\hat{\gamma}_{bctm}^*$ | $\hat{\gamma}_{swcqr}^*$ | $\hat{\gamma}_{ntm}^*$ |
|---------|----------|-------------------|-----------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 3a      | 0        | Monthly           | 52.997                | 52.790               | 52.815                | 52.710                | 52.740                | 60.786                | 52.701                |
|         |          | Daily             | 52.797                | 52.032               | 52.721                | 62.049                |                       |                       |                       |
|         |          | Hourly            | 52.846                | 52.148               | 52.762                | 63.837                |                       |                       |                       |
|         | 200      | Monthly           | 52.881                | 52.651               | 52.626                | 53.189                | 53.198                | 60.427                | 54.940                |
|         |          | Daily             | 52.662                | 52.378               | 53.251                | 61.442                |                       |                       |                       |
|         |          | Hourly            | 52.740                | 62.024               | 53.393                | 64.471                |                       |                       |                       |
|         | 300      | Monthly           | 52.996                | 52.692               | 53.083                | 53.743                | 53.889                | 62.495                | 53.716                |
|         |          | Daily             | 52.841                | 52.930               | 53.743                | 62.495                |                       |                       |                       |
|         |          | Hourly            | 52.794                | 62.161               | 54.220                | 58.452                |                       |                       |                       |
|         | 500      | Monthly           | 52.887                | 52.745               | 54.375                | 55.447                | 56.326                | 64.413                | 59.138                |
|         |          | Daily             | 52.839                | 53.238               | 54.465                | 63.881                |                       |                       |                       |
|         |          | Hourly            | 52.912                | 65.057               | 57.030                | 69.825                |                       |                       |                       |
|         | 800      | Monthly           | 52.701                | 52.554               | 54.671                | 59.812                | 60.117                | 61.720                | 61.594                |
|         |          | Daily             | 53.149                | 67.921               | 62.105                | 71.426                |                       |                       |                       |
|         |          | Hourly            | 52.882                | 69.403               | 59.779                | 62.084                |                       |                       |                       |
| 3b      | 0        | Monthly           | 91.742                | 91.462               | 91.397                | 91.359                | 91.375                | 95.290                | 91.446                |
|         |          | Daily             | 91.501                | 106.725              | 91.408                | 104.471               |                       |                       |                       |
|         |          | Hourly            | 91.780                | 126.465              | 91.364                | 118.268               |                       |                       |                       |
|         | 200      | Monthly           | 91.565                | 91.482               | 91.708                | 91.979                | 91.993                | 97.766                | 96.342                |
|         |          | Daily             | 91.474                | 103.885              | 91.991                | 98.795                |                       |                       |                       |
|         |          | Hourly            | 91.496                | 126.705              | 92.030                | 117.557               |                       |                       |                       |
|         | 300      | Monthly           | 91.584                | 91.474               | 92.273                | 92.586                | 92.654                | 97.938                | 98.314                |
|         |          | Daily             | 91.459                | 158.844              | 92.612                | 142.741               |                       |                       |                       |
|         |          | Hourly            | 91.662                | 173.762              | 92.724                | 159.019               |                       |                       |                       |
|         | 500      | Monthly           | 91.631                | 91.480               | 94.211                | 95.129                | 94.452                | 94.651                | 105.598               |
|         |          | Daily             | 92.959                | 152.931              | 95.111                | 144.314               |                       |                       |                       |
|         |          | Hourly            | 91.451                | 159.302              | 95.065                | 151.774               |                       |                       |                       |
|         | 800      | Monthly           | 91.571                | 91.428               | 96.029                | 98.257                | 99.787                | 100.455               | 102.837               |
|         |          | Daily             | 91.989                | 162.616              | 99.593                | 150.618               |                       |                       |                       |
|         |          | Hourly            | 91.429                | 187.443              | 98.438                | 166.571               |                       |                       |                       |

### Table 7 The MAEs of various estimators in fitting the data in the test set from the Beijing Multi-Site Air-Quality Data set

| Example | $\gamma$ | Data chunk levels | $\hat{\gamma}_{kcqr}^*$ | $\hat{\gamma}_{ntm}^*$ | $\hat{\gamma}_{swcqr}^*$ | $\hat{\gamma}_{bctm}^*$ | $\hat{\gamma}_{bctm}^*$ | $\hat{\gamma}_{swcqr}^*$ | $\hat{\gamma}_{ntm}^*$ |
|---------|----------|-------------------|-----------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 3a      | 0        | Monthly           | 39.917                | 40.205               | 40.631                | 40.508                | 40.512                | 40.671                | 40.580                |
|         |          | Daily             | 40.239                | 44.321               | 40.506                | 43.978                |                       |                       |                       |
|         |          | Hourly            | 40.282                | 46.134               | 40.549                | 43.749                |                       |                       |                       |
|         | 200      | Monthly           | 40.049                | 40.357               | 39.746                | 39.824                | 39.835                | 43.451                | 41.376                |
|         |          | Daily             | 40.363                | 44.904               | 39.801                | 43.793                |                       |                       |                       |
|         |          | Hourly            | 40.415                | 46.022               | 39.940                | 44.959                |                       |                       |                       |
|         | 300      | Monthly           | 40.150                | 40.348               | 39.724                | 39.658                | 39.805                | 43.378                | 41.964                |
|         |          | Daily             | 40.588                | 45.810               | 39.662                | 42.255                |                       |                       |                       |
|         |          | Hourly            | 40.502                | 46.724               | 39.999                | 44.666                |                       |                       |                       |
|         | 500      | Monthly           | 40.063                | 40.419               | 39.596                | 39.784                | 39.801                | 42.366                | 44.366                |
|         |          | Daily             | 40.547                | 43.735               | 40.453                | 51.699                |                       |                       |                       |
|         |          | Hourly            | 40.600                | 54.121               | 41.430                | 55.619                |                       |                       |                       |
|         | 800      | Monthly           | 39.886                | 40.323               | 39.460                | 43.198                | 44.917                | 44.878                | 46.053                |
data, the errors of \( \hat{\gamma}_{\text{bctm}}, \hat{\gamma}_{\text{swcqr}} \) and \( \hat{\gamma}_{\text{nw}} \) are close to each other. While, as \( \gamma \) is increasing, the RMSEs of \( \hat{\gamma}_{\text{bctm}} \) and \( \hat{\gamma}_{\text{swcqr}} \) are enlarged but their MAEs change little. For \( \gamma > 300 \), both of \( \hat{\gamma}_{\text{bctm}} \) and \( \hat{\gamma}_{\text{swcqr}} \) are significantly superior than \( \hat{\gamma}_{\text{nw}} \) in terms of RMSEs and MAEs. The above results show that our renewable BCTM enjoy robustness stronger than the NW estimators.

In summary, by comprehensively investigating the numerical results in various experiment conditions, we can conclude the desirable performance of our estimation method and algorithms.

### 6 Conclusions

In the previous sections, we proposed a new approach to constructing composite quantile estimators for the nonparametric model (1) with streaming data.

A new feature in our method is that the nonparametric function of interest is expressed as a functional that depends on a weight function and a conditional distribution function. Then the CDF is estimated by renewable kernel estimations together with function interpolations. Based on the renewable estimators, we proposed the method of renewable WCQR estimation for our target function.

By the model structure, we found new selectors for the weight function, under which the WCQR can achieve asymptotic unbiasedness when estimating regression function \( m(\cdot) \) and variance function \( \sigma(\cdot) \) in the model (1). Then we suggested the renewable NTM and BCTM to estimate \( m(\cdot) \), and the renewable NTSD and RTSD to estimate \( \sigma(\cdot) \). Our theoretical analyses and numerical experiments demonstrate that the proposed renewable estimators all enjoy desirable performance comparable to the oracle estimators computed on the full data sets. Moreover, they are more computational efficient than other optimization-based estimators.

Based on the theoretical properties and simulation experiments, we have the following recommendations for the uses of the proposed estimators:

- For symmetric model (1), the NTM is a consistent and robust estimator for \( m(\cdot) \). But when the model (1) is asymmetric, the NTM suffers from non-negligible bias. Thus the NTM is recommended for symmetric models with possibly contaminated streaming data.
- The BCTM is always a consistent estimator for \( m(\cdot) \) regardless of whether the model (1) is symmetric or not, but its robustness is inferior than the NTM; see Remark 3.6. Thus the BCTM is recommended when the model is asymmetric or the symmetry of the error distribution is unknown.
- The NTSD and RTSD are both estimators for the variance function \( \sigma(\cdot) \). The NTSD enjoys robustness but it only estimates \( \sigma(x) \) up to scale factor \( c_0 \), i.e., it cannot completely identify the unknown constant \( c_0 \) in (36). The RTSD can identify \( c_0 \) and enjoys consistency for \( \sigma(x) \), but its robustness is inferior than the NTSD; see Remark 3.7. Thus the NTSD is recommended if there is no need for identifying the unknown \( c_0 \), otherwise the RTSD is recommended.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11222-023-10352-x.
Funding The research was supported by NNSF project of China (11971265) and National Key R &D Program of China (2018YFA0703900).

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Ashfahani, A., Pratama, M.: Autonomous Deep Learning: Continual Learning Approach for Dynamic Environments, pp. 666–674 (2019)

Bednar, J., Watt, T.: Alpha-trimmed means and their relationship to median filters. IEEE Trans. Acoust. Speech Signal Process. 32(1), 145–153 (1984). https://doi.org/10.1109/TASSP.1984.1164279

Bickel, P.J., Lehmann, E.L.: Descriptive statistics for nonparametric models. II. Location. Ann. Stat. 3(5), 1045–1069 (1975)

Boente, G., Fraiman, R.: Local L-estimators for nonparametric regression under dependence. J. Nonparametr. Statist. 4(1), 91–101 (1994). https://doi.org/10.1080/10485259408832603

Bucak, S.S., Gunsel, B.: Incremental subspace learning via non-negative matrix factorization. Pattern Recogn. 42(5), 788–797 (2009). https://doi.org/10.1016/j.patcog.2008.09.002

Burden, R.L., Faires, J.D., Burden, A.M.: Numerical Analysis. Cengage Learning (2015)

Chen, X., Liu, W., Zhang, Y.: Quantile regression under memory constraint. Ann. Stat. 47(6), 3244–3273 (2019). https://doi.org/10.1214/18-AOS1777

Das, M., Pratama, M., Savitri, S., Zhang, J.: Multilayer self-evolving recurrent neural network for data stream classification. In: 2019 IEEE International Conference on Data Mining (ICDM), pp. 110–119 (2019)

Fan, J.: Local linear regression smoothers and their minimax efficiencies. Ann. Stat. 21(1), 196–216 (1993). https://doi.org/10.1214/aos/1176349022

Fan, J., Gijbels, I.: Variable bandwidth and local linear regression smoothers. Ann. Stat. 20(4), 2008–2036 (1992). https://doi.org/10.1214/aos/1040118511

Fu, Y., Zhao, W., Zhou, T.: Efficient spectral sparse grid approximations for solving multi-dimensional forward backward SDEs. Discret. Contin. Dyn. Syst. Ser. B 22(9), 3439–3458 (2017). https://doi.org/10.3934/dcdsb.20171714

Gautschi, W.: Numerical Analysis, 2nd edn. Birkhäuser, Boston, MA (2012)

Gutenbrunner, C., Jurečková, J.: Regression rank scores and regression quantiles. Ann. Stat. 20(1), 305–330 (1992). https://doi.org/10.1214/aos/1176348524

Jiang, R., Qian, W.M., Zhou, Z.G.: Single-index composite quantile regression with heteroscedasticity and general error distributions. Stat. Pap. 57(1), 185–203 (2016). https://doi.org/10.1007/s00362-014-0646-y

Kai, B., Li, R., Zou, H.: Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. J. R. Stat. Soc. Ser. B Stat Methodol. 72(1), 49–69 (2010). https://doi.org/10.1111/j.1467-9868.2009.00725.x

Koenker, R.: Quantile Regression, Volume 38 of Econometric Society Monographs. Cambridge University Press, Cambridge (2005)

Koenker, R., Portnoy, S.: L-estimation for linear models. J. Am. Stat. Assoc. 82(399), 851–857 (1987)

Koenker, R., Zhao, Q.S.: L-estimation for linear heteroscedastic models. J. Nonparametr. Statist. 3(3–4), 223–235 (1994). https://doi.org/10.1080/10485259408832584

Lin, L., Li, F., Wang, K., Zhu, L.: Composite estimation: an asymptotically weighted least squares approach. Stat. Sinica 29(3), 1367–1393 (2019)

Lin, L., Li, W., Lu, J.: Unified rules of weighted least squares for various online updating estimations. arXiv:2008.08824 (2020)

Luo, L., Song, P.X.K.: Adaptive L-estimation and incremental inference in generalized linear models with streaming data sets. J. R. Stat. Soc. Ser. B Stat Methodol. 82(1), 69–97 (2020)

Moroshko, E., Vaits, N., Crammer, K.: Second-order non-stationary online learning for regression. J. Mach. Learn. Res. 16, 1481–1517 (2015)

Nion, D., Sidiroopoulos, N.D.: Adaptive algorithms to track the parafac decomposition of a third-order tensor. IEEE Trans. Signal Process. 57(6), 2299–2310 (2009). https://doi.org/10.1109/TSP.2009.2016885

Portnoy, S., Koenker, R.: Adaptive L-estimation for linear models. Ann. Stat. 17(1), 362–381 (1989). https://doi.org/10.1214/aos/1176348900

Ruppert, D., Wand, M.P., Holst, U., Hössjer, O.: Local polynomial variance-function estimation. Technometrics 39(3), 262–273 (1997)

Sauer, T.: Numerical Analysis. Addison-Wesley Publishing Company (2011)

Scharf, E.D., Wu, J., Wang, C., Yan, J., Chen, M.H.: Online updating of statistical inference in the big data setting. Technometrics 58(3), 393–403 (2016). https://doi.org/10.1080/00401706.2016.1142900

Serfling, R.J.: Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York (1980)

Shang, Z., Cheng, G.: Computational limits of a distributed algorithm for smoothing spline. J. Mach. Learn. Res. 18, 37 (2017)

Stigler, S.M.: Do robust estimators work with real data? Ann. Stat. 5(6), 1055–1098 (1977)

Sun, J., Gai, Y., Lin, L.: Weighted local linear composite quantile estimation for the case of general error distributions. J. Stat. Plann. Inference 143(6), 1049–1063 (2013). https://doi.org/10.1016/j.jspi.2013.01.002

Toulis, P., Rennie, J., Airoldi, E.: Statistical analysis of stochastic gradient methods for generalized linear models. Int. Conf. Mach. Learn. 32(1), 667–675 (2014)

Volgushev, S., Chao, S.K., Cheng, G.: Distributed inference for quantile regression processes. Ann. Stat. 47(3), 1634–1662 (2019). https://doi.org/10.1214/18-AOS1730

Wang, K., Li, S., Zhang, B.: Robust communication-efficient distributed composite quantile regression and variable selection for massive data. Comput. Stat. Data Anal. 161, 107262 (2021). https://doi.org/10.1016/j.csda.2021.107262

Wang, K., Wang, H., Li, S.: Renewable quantile regression for streaming datasets. Knowl.-Based Syst. 235, 107675 (2022). https://doi.org/10.1016/j.knosys.2021.107675

Zhao, W., Chen, L., Peng, S.: A new kind of accurate numerical method for backward stochastic differential equations. SIAM J. Sci. Comput. 28(4), 1563–1581 (2006). https://doi.org/10.1137/ 05063341X

Zhao, W., Fu, Y., Zhou, T.: New kinds of high-order multistep schemes for coupled forward backward stochastic differential equations.
Zou, H., Yuan, M.: Composite quantile regression and the oracle model selection theory. Ann. Stat. 36(3), 1108–1126 (2008). https://doi.org/10.1214/07-AOS507

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g., a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.