JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. (2023)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12725

ORIGINAL ARTICLE

NON-CROSSING QUANTILE DOUBLE-AUTOREGRESSION FOR THE ANALYSIS OF STREAMING TIME SERIES DATA

RONG JIANG[a] SIU KAI CHOY[b] AND KEMING YU[c,d*]

[a]*School of Mathematics Physics and Statistics, Shanghai Polytechnic University, Shanghai, People's Republic of China*
[b]*Department of Mathematics, Statistics and Insurance, The Hang Seng University of Hong Kong, Hong Kong*
[c]*Department of mathematics, Brunel University London, UK*
[d]*College of Mathematics and Physics, Anqing Normal University, Anqing, People's Republic of China*

Many financial time series not only have varying structures at different quantile levels and exhibit the phenomenon of conditional heteroscedasticity at the same time but also arrive in the stream. Quantile double-autoregression is very useful for time series analysis but faces challenges with model fitting of streaming data sets when estimating other quantiles in subsequent batches. This article proposes a renewable estimation method for quantile double-autoregression analysis of streaming time series data due to its ability to break with storage barrier and computational barrier. Moreover, the proposed flexible parametric structure of the quantile function enables us to predict any interested quantile value without quantile curve crossing problem or keeping the desirable monotone property of the conditional quantile function. The proposed methods are illustrated using current data and the summary statistics of historical data. Theoretically, the proposed statistic is shown to have the same asymptotic distribution as the standard version computed on an entire data stream with the data batches pooled into one data set, without additional condition. Simulation studies and an empirical example are presented to illustrate the finite sample performance of the proposed methods.

Received 23 February 2023; Accepted 19 September 2023

Keywords: Double autoregressive model; generalized lambda distribution; online updating; quantile regression; quantile curve crossing; streaming data.

MOS subject classification: 60G08; 62G20.

1. INTRODUCTION

In economics and finance, considerable attention has been devoted to regression models with autoregressive errors for time series data. Because volatility is fundamental to asset pricing, monetary policymaking, portfolio management and risk analysis, it is especially important to accurately forecast volatility. To capture the time-varying volatility, it is necessary to take the conditional heteroscedasticity into account when a linear model is fitted to financial time series. Among existing conditional heteroscedastic models, the double-autoregressive (DAR) models (Ling, 2004) have recently attracted growing attention, see Cai *et al.* (2013), Li *et al.* (2015), Xu and Zhao (2021), Zhu and Li (2022) and among others. Because it has two novel properties. First, it has a larger parameter space than that of the commonly used autoregressive (AR) model. Second, the quasi-maximum-likelihood estimator for the DAR model is still asymptotically normal without assuming the moment condition on returns (Ling, 2007), which does not hold for classic AR(p) model with independent and identically distributed errors. The DAR model of order p is defined as

$$Y_t = \boldsymbol{X}_t^{\top}\boldsymbol{\beta}_0 + \sigma(\boldsymbol{X}_t, \boldsymbol{\alpha}_0)\epsilon_t, \quad t = 1, \ldots, N, \tag{1.1}$$

* Correspondence to: Keming Yu, Department of Matheamtics, Brunel University London, Kingston Lane, Uxbridge, UB9 3PH, UK.
Email: keming.yu@brunel.ac.uk

where $\boldsymbol{\beta}_0$ and $\boldsymbol{\alpha}_0$ are vector of unknown parameters, the noise ϵ_t is independent of p-dimensional covariates $\boldsymbol{X}_t = (Y_{t-1}, \ldots, Y_{t-p})^\top$ and $\sigma(\boldsymbol{X}_t, \boldsymbol{\alpha}_0) = \sqrt{1 + \sum_{j=1}^{p} \alpha_{0,j} Y_{t-j}^2}$.

Modeling conditional quantiles and volatility dynamics together is of extreme importance in econometrics and finance. Obviously, the maximum likelihood and least squares estimations are both lack of robustness and may cause larger bias. The quantile regression (QR) method proposed by Koenker and Bassett (1978) can not only maintain the robustness of estimation, but also capture the characteristics of the whole conditional distribution. This allows us to estimate the upper or lower tail of the conditional distribution of interest. In particular, it has been widely used for the prediction of quantile-based risk measures, for example, the value at risk. Hence, it is natural to consider QR for the DAR model as

$$Q_{Y_t|\mathcal{F}_{t-1}}(\tau) = \boldsymbol{X}_t^\top \boldsymbol{\beta}_0 + \sigma(\boldsymbol{X}_t, \boldsymbol{\alpha}_0) Q_{\epsilon_t}(\tau), \tag{1.2}$$

where $Q_{Y_t|\mathcal{F}_{t-1}}(\tau)$ is the conditional τth quantile of Y_t given the up to $t-1$ information set $\mathcal{F}_{t-1}$ and $Q_{\epsilon_t}(\tau)$ is the τth quantile of ϵ_t. For model (1.2), Cai *et al.* (2013) developed a Bayesian estimation method, Xu and Zhao (2021) proposed an efficient estimator for $\boldsymbol{\beta}_0$ by constrainedly weighting information across quantiles, and Zhu and Li (2022) studied a self-weighted conditional quantile estimation method to estimate $(\boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0^\top, Q_{\epsilon_t}(\tau))^\top$ by minimizing the following function:

$$\operatorname{argmin}_{(\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, b_\tau)^\top} \sum_{t=p+1}^{N} \omega_t \rho_\tau \left(Y_t - \boldsymbol{X}_t^\top \boldsymbol{\beta} + \sigma(\boldsymbol{X}_t, \boldsymbol{\alpha}) b_\tau \right), \tag{1.3}$$

where $\rho_\tau(r) = \tau r - r\mathrm{I}(r < 0)$ is check loss function, $\mathrm{I}(\cdot)$ is the indicator function and $\{\omega_t\}_{t=p+1}^{N}$ are non-negative random weights.

Our era has witnessed the massive explosion of data and a dramatic improvement on technology in collecting and processing big data. Due to the explosive growth of data onto non-traditional sources such as mobile phones, social networks, and e-commerce, streaming data are becoming a core component in big data analysis. For example, when a passenger calls Lyft, real-time streams of data join together to create a seamless user experience. Through this data, the application pieces together real-time location tracking, traffic stats, pricing, and real-time traffic data to simultaneously match the rider with the best possible driver, calculate pricing, and estimate time to destination based on both real-time and historical data. As streaming data grows rapidly in volume and velocity, storing and combing data becomes increasingly challenging due to limited computer memory and storage. To reduce the demand on computing memory and achieve real-time processing, the nature of streaming data calls for the development of algorithms which require only 'one pass' over the data. However, the ordinary QR estimator does not have a display expression, because the loss function $\rho(\cdot)$ in (1.3) is not a convex function. As a result, if a quantile τ is not considered at the first batch, we cannot obtain its estimator in the subsequent batches, as in ordinary quantile regression, because the data is one pass. Moreover, it is impossible to estimate all quantiles τ in $(0, 1)$ at the beginning, which also increases the computational burden. In this article, we address the following natural question:

> Can we construct a QR estimator of $Q_{Y_t|\mathcal{F}_{t-1}}(\tau)$ in the model (1.2) with an explicit expression, which is an increasing function of quantile level τ?

For estimating the unknown parameters in the classical regression models under streaming data, Schifano *et al.* (2016) developed online-updating algorithms for linear models and estimating equations. Luo and Song (2020) proposed a renewable estimation for the generalized linear model. Luo *et al.* (2022) developed an incremental learning algorithm based on quadratic inference function to analyze streaming datasets with correlated outcomes such as longitudinal data and clustered data. Yang and Yao (2022) studied an online non-parametric method to dynamically update the estimates of mean and covariance functions for functional data.

DOI: 10.1111/jtsa.12725

Quan and Lin (2022) considered a one-pass non-parametric estimation method for non-parametric regression in the streaming setting. The above methods for streaming data are all based on ordinary least squares or estimating equations. Due to the second-order non-differentiable of the loss function $\rho(\cdot)$ in (1.3), we find that the above methods for streaming data based on the least squares and estimating equations are not suitable for QR estimator. To overcome the non-differentiability of the QR loss function, Jiang and Yu (2022) used a convolution-type smoothing method to develop a renewable estimation. Chen *et al.* (2019) and Wang *et al.* (2022) also studied QR estimation for streaming data. However, their methods are all required additional strict conditions on the sample size of each bath. The above three QR methods for streaming data analysis, like (1.3), do not solve the problem presented, because they still can only obtain an estimator at a quantile level at a time.

Note that model (1.2), only $Q_{\varepsilon_t}(\tau)$ depends on τ, while $\boldsymbol{\beta}_0$ and $\boldsymbol{\alpha}_0$ are independent of τ. We adopt a parameterized method to estimate $Q_{\varepsilon_t}(\tau)$, then a QR estimator of $Q_{Y_t|\mathcal{F}_{t-1}}(\tau)$ can be obtained, which is a function of quantile level τ. The easiest way to think of parameterizing $Q_{\varepsilon_t}(\tau)$ is $\boldsymbol{\theta} b(\tau)$ (Frumento and Bottai, 2016), where $b(\tau)$ is a set of known functions of τ. For instance, $Q_{\varepsilon_t}(\tau) = \theta_0 + \theta_1\tau$ under $b(\tau) = (1, \tau)^\top$. However, it is difficult to parameterize the quantile regression coefficients accurately, and their method cannot guarantee obtain the non-crossing of different quantile estimators (Frumento *et al.*, 2021). Population conditional quantile functions cannot cross each other for different quantile orders; however, the estimated regression curves often violate this (Jiang and Yu, 2023). For example, in some cases, the estimate of the 5th percentile is larger than that of the 10th percentile, which can be very challenging for interpretation and further analysis.

In this article, we parameterized $Q_{\varepsilon_t}(\tau)$ as function (1.4) by the generalized lambda distribution (GLD), because the GLD (Freimer *et al.*, 1988) is defined by its quantile function (the inverse of the distribution function).

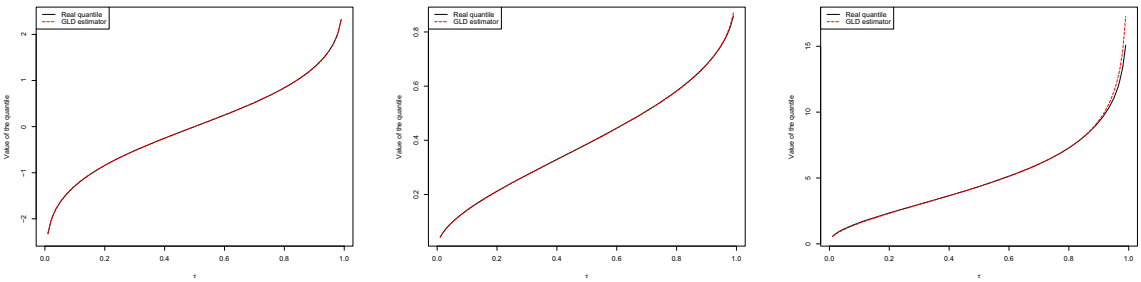
$$Q_{\varepsilon_t}(\tau) = \theta_{0,1} + \theta_{0,2}\left\{\frac{\tau^{\theta_{0,3}} - 1}{\theta_{0,3}} - \frac{(1-\tau)^{\theta_{0,4}} - 1}{\theta_{0,4}}\right\}, \tag{1.4}$$

where $\theta_{0,1}$ is a location parameter, $\theta_{0,2} > 0$ an inverse scale parameter and $\theta_{0,3}$, $\theta_{0,4}$ are shape parameters. The procedure (1.4) allows to estimate the whole quantile function of ε_t directly using a wider class of distributions, including those which are defined only via their quantile functions and that may not have closed mathematical expressions of their density or distribution functions. Given the correct parameters, the GLD distribution equates to several well-known distributions (e.g. the uniform, exponential, logistic) and for many others, a close approximation is possible (e.g. Gaussian, Cauchy, Student's t, chi-square, Gamma, Weibull, lognormal, beta distribution). Figure 1 illustrates that GLD estimates the quantile function well, where the parameter of GLD is obtained by Dedduwakumara *et al.* (2021). We refer the reader to Karian and Dudewicz (2000) for a complete list of distributions that the GLD can represent and their corresponding parameters. Cai *et al.* (2013), Cai (2016) and Cai and Li (2019) considered Bayesian estimations for double AR time series models, nonlinear time series models and threshold GARCH models, respectively, with ε_t following a generalized lambda distribution. Zhang *et al.* (2021) developed a quantile index regression by Tukey lambda distribution, and their proposed estimation method is not suitable for heteroscedasticity time series data. Combing (1.2) and (1.4), we have

$$Q_{Y_t|\mathcal{F}_{t-1}}(\tau) = \boldsymbol{X}_t^\top\boldsymbol{\beta}_0 + \sigma(\boldsymbol{X}_t, \boldsymbol{\alpha}_0)\left[\theta_{0,1} + \theta_{0,2}\left\{\frac{\tau^{\theta_{0,3}} - 1}{\theta_{0,3}} - \frac{(1-\tau)^{\theta_{0,4}} - 1}{\theta_{0,4}}\right\}\right], \tag{1.5}$$

where the unknown parameters $\boldsymbol{\gamma}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0^\top, \theta_{0,1}, \theta_{0,2}, \theta_{0,3}, \theta_{0,4})^\top$ are all independent of τ. The approach (1.5) allows us to estimate the quantile function of a response variable, rather than a sequence of quantiles. Moreover, it is worth noting that we can obtain non-crossing quantile estimators of $Q_{Y_t|\mathcal{F}_{t-1}}(\tau)$ under different τ according to

$$\partial Q_{Y_t|\mathcal{F}_{t-1}}(\tau)/\partial\tau = \sigma(\boldsymbol{X}_t, \boldsymbol{\alpha}_0)\theta_{0,2}\left\{\tau^{\theta_{0,3}-1} + (1-\tau)^{\theta_{0,4}-1}\right\} > 0,$$



(a) N(0,1) (b) Beta(2,3) (c) $\chi^2(5)$

Figure 1. GLD estimation for quantile functions of three distributions

where $\sigma(\boldsymbol{X}_t, \boldsymbol{\alpha}_0)$ and $\theta_{0,2}$ are all positive. Model (1.5) also enables us to conduct the estimation of quantile levels with rich observations and then to extrapolate the fitted structures to far tail, such as Value-at-Risk in economics and finance.

To summarize, we develop a QR method for the DAR model (1.1) with streaming datasets. We first parameterize $Q_{\varepsilon_t}(\tau)$ as function (1.4) by the generalized lambda distribution (GLD). The unknown parameters $\boldsymbol{\gamma}_0$ in the model (1.5) can be estimated by a convolution-type smoothing method, which only requires the availability of the current data batch in the data stream and sufficient statistics of the historical data at each stage of the analysis. Theoretically, it has the same asymptotic distributions as the standard version computed on an entire data stream with the data batches pooled into one data set, without additional condition. Based on (1.5), we can estimate the conditional quantile at a very high or low level of quantile level, and the quantile estimators of $Q_{Y_t|\mathcal{F}_{t-1}}(\tau)$ under different τs are non-crossing. The most important thing is that it can be applied to streaming data analysis to obtain any wanted conditional quantile estimation.

The remainder of this article is organized as follows. In Section 2, the estimation methods based on all data are proposed. The renewable estimation method is developed in Section 3. Both simulation examples and the application of real data is given in Section 4 to illustrate the proposed procedures. We conclude this article with a brief discussion in Section 5. All technical proofs are provided in the Appendix.

2. ESTIMATION METHOD BASED ON ALL DATA

2.1. Self-weighted Composite Quantile Regression Estimation

We first study the estimation method based on all data with sample size N. Note that the unknown parameters $\boldsymbol{\gamma}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0^\top, \theta_{0,1}, \theta_{0,2}, \theta_{0,3}, \theta_{0,4})^\top$ are all independent of τ, we therefore consider the following self-weighted composite quantile regression (SWCQR) to achieve high efficiency:

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \rho_{\tau_k} \left(Y_t - q_t(\boldsymbol{\gamma}, \tau_k) \right), \tag{2.1}$$

where K is a fixed integer with $0 < \tau_1 < \cdots < \tau_K < 1$ and one can use the equally spaced quantiles at $\tau_k = k/(K+1)$ for $k = 1, \ldots, K$ (Zou and Yuan, 2008), $q_t(\boldsymbol{\gamma}, \tau) = \boldsymbol{X}_t^\top \boldsymbol{\beta} + \sigma(\boldsymbol{X}_t, \boldsymbol{\alpha}) \left(\theta_1 + \theta_2 \left[(\tau^{\theta_3} - 1)/\theta_3 - \{(1-\tau)^{\theta_4} - 1\}/\theta_4 \right] \right)$, $\boldsymbol{\gamma} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \theta_1, \theta_2, \theta_3, \theta_4)^\top$ and $\{\omega_t\}_{t=1}^N$ are non-negative random weights, which are to maintain the asymptotic normality for heavy-tailed data with only a finite fractional moment, see Ling (2005) and Zhu and Li (2022).

Therefore, we can obtain the estimator $\hat{Q}_{Y_t|\mathcal{F}_{t-1}}(\tau)$ of $Q_{Y_t|\mathcal{F}_{t-1}}(\tau)$ based on $\hat{\boldsymbol{\gamma}}$ as

$$\hat{Q}_{Y_t|\mathcal{F}_{t-1}}(\tau) = q_t(\hat{\boldsymbol{\gamma}}, \tau) = \boldsymbol{X}_t^\top \hat{\boldsymbol{\beta}} + \sigma(\boldsymbol{X}_t, \hat{\boldsymbol{\alpha}}) \left[\hat{\theta}_1 + \hat{\theta}_2 \left\{ \frac{\tau^{\hat{\theta}_3} - 1}{\hat{\theta}_3} - \frac{(1-\tau)^{\hat{\theta}_4} - 1}{\hat{\theta}_4} \right\} \right]. \tag{2.2}$$

DOI: 10.1111/jtsa.12725

Note that $\hat{Q}_{Y_t|\mathcal{F}_{t-1}}(\tau)$ in (2.2) is a function of τ, thus we can use it to estimate all conditional quantile functions, especially for extreme conditional quantile functions. Moreover, $\hat{Q}_{Y_t|\mathcal{F}_{t-1}}(\tau)$ under different $\tau \in (0,1)$ are non-crossing by

$$\partial \hat{Q}_{Y_t|\mathcal{F}_{t-1}}(\tau)/\partial\tau = \sigma(\boldsymbol{X}_t, \hat{\boldsymbol{\alpha}})\hat{\theta}_2 \left\{ \tau^{\hat{\theta}_3 - 1} + (1-\tau)^{\hat{\theta}_4 - 1} \right\} > 0,$$

where $\sigma(\boldsymbol{X}_t, \hat{\boldsymbol{\alpha}})$ and $\hat{\theta}_2$ are all positive.

2.2. Smoothing Self-weighted Composite Quantile Estimation

Because the loss function $\rho_\tau(r) = \tau r - r\mathrm{I}(r < 0)$ in (1.3) is non-differentiable, the SWCQR estimator in (2.1) has no display expression, it is impossible to construct a renewable estimator for streaming data sets. To circumvent the non-differentiability of the QR loss function, we estimate $\boldsymbol{\gamma}_0$ by minimizing the following smoothing quantile regression objective function:

$$\begin{aligned} \tilde{\boldsymbol{\gamma}}^* &= \arg\min_{\boldsymbol{\gamma}} S_h(\boldsymbol{\gamma}) \\ &= \arg\min_{\boldsymbol{\gamma}} \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \int_{-\infty}^{+\infty} \rho_{\tau_k}(s) K_h(s - Y_t + q_t(\boldsymbol{\gamma}, \tau_k)) \mathrm{d}s, \end{aligned} \tag{2.3}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a smooth kernel function and h is a bandwidth. The kernel smooth function is used in (2.3) to estimate the distribution function, and the indicator function is used in (2.1), the details can see Fernandes *et al.* (2021). Now $S_h(\boldsymbol{\gamma})$ is twice continuously differentiable with the gradient and Hessian matrix

$$\begin{aligned} \nabla S_h(\boldsymbol{\gamma}) &= \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \nabla q_t(\boldsymbol{\gamma}, \tau_k) \left\{ K^*((q_t(\boldsymbol{\gamma}, \tau_k) - Y_t)/h) - \tau_k \right\}, \\ \nabla^2 S_h(\boldsymbol{\gamma}) &= \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \Big[\nabla^2 q_t(\boldsymbol{\gamma}, \tau_k) \left\{ K^*((q_t(\boldsymbol{\gamma}, \tau_k) - Y_t)/h) - \tau_k \right\} \\ &\quad + \nabla q_t(\boldsymbol{\gamma}, \tau_k) \left\{ \nabla q_t(\boldsymbol{\gamma}, \tau_k) \right\}^\top K_h(q_t(\boldsymbol{\gamma}, \tau_k) - Y_t) \Big], \end{aligned} \tag{2.4}$$

respectively, where $K^*(v) = \int_{-\infty}^{v} K(u)\mathrm{d}u$, ∇ and ∇^2 are the first and second derivatives of $\boldsymbol{\gamma}$, respectively. For example, we can take an Epanechnikov kernel $K(u) = (3/4)(1-u^2)\mathrm{I}(|u| \le 1)$, then

$$\begin{aligned} S_h(\boldsymbol{\gamma}) = \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \Big[&\frac{h}{16} \{3 + 6e_{tk}^2(\boldsymbol{\gamma})/h^2 - e_{tk}^4(\boldsymbol{\gamma})/h^4\} \mathrm{I}\{|e_{tk}(\boldsymbol{\gamma})| \le h\} \\ &+ \frac{1}{2}|e_{tk}(\boldsymbol{\gamma})| \mathrm{I}\{|e_{tk}(\boldsymbol{\gamma})| > h\} + (\tau - 1/2) e_{tk}(\boldsymbol{\gamma}) \Big], \end{aligned}$$

where $e_{tk}(\boldsymbol{\gamma}) = Y_t - q_t(\boldsymbol{\gamma}, \tau_k)$.

2.3. Large Sample Properties

To establish the asymptotic properties of the proposed estimators, the following technical conditions are imposed.

C1. $\{Y_t\}$ is strictly stationary and ergodic with $\mathrm{E}|Y_t|^r < \infty$ for some $0 < r \leq 1$. The conditional density function of Y_t on $\mathcal{F}_{t-1}$ ($f_{Y_t|\mathcal{F}_{t-1}}$) and its derivative function ($f'_{Y_t|\mathcal{F}_{t-1}}$) are uniformly bounded and $f_{Y_t|\mathcal{F}_{t-1}}$ is positive on the support $\{x : 0 < F_{Y_t|\mathcal{F}_{t-1}}(x) < 1\}$, where $F_{Y_t|\mathcal{F}_{t-1}}$ is the distribution function of Y_t on $\mathcal{F}_{t-1}$.

C2. $\{\omega_t\}$ is strictly stationary, ergodic, non-negative and measurable with respect to $\boldsymbol{X}_t$ with $\mathrm{E}(\omega_t \|\boldsymbol{X}_t\|_2^3) < \infty$, where $\|\cdot\|_2$ is the 2-norm.

C3. $\boldsymbol{\Sigma}_1 = \sum_{k=1}^{K} \mathrm{E}[f_{Y_t|\mathcal{F}_{t-1}}(q_t(\boldsymbol{\gamma}_0, \tau_k))\omega_t \nabla q_t(\boldsymbol{\gamma}_0, \tau_k)\{\nabla q_t(\boldsymbol{\gamma}_0, \tau_k)\}^\top]$ is a positive definite matrix.

C4. The kernel function $K(\cdot)$ is even, integrable, and twice differentiable with bounded first and second derivatives such that $\int K(u)\mathrm{d}u = 1$, $\int |u^2 K(u)|\mathrm{d}u < \infty$, $\int uK(u)\mathrm{d}u = 0$ and $\int u^2 K(u)\mathrm{d}u \neq 0$. In addition, $0 < \int_0^\infty K^*(u)\{1 - K^*(u)\}\mathrm{d}u < \infty$.

Remark 2.1. Condition **C1** is commonly used in the literature, see Zhu and Li (2022). For the random weights $\{\omega_t\}$, there are many choices satisfying Condition **C2**, such as $\omega_t = (1 + \sum_{j=1}^{p} |Y_{t-j}|^3)^{-1}$. Condition **C3** ensures that $\boldsymbol{\Sigma}_1^{-1}$ exists. Condition **C4** is a mild condition on $K(\cdot)$ for smoothing approximation. For example, Epanechnikov kernel $K(u) = (3/4)(1 - u^2)I(|u| \leq 1)$ satisfies condition **C4**.

Theorem 2.1. Suppose that conditions **C1**–**C3** are satisfied and $N \to \infty$. Then, we have

$$\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{L} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\right),$$

and

$$\hat{Q}_{Y_{N+1}|\mathcal{F}_N}(\tau) - Q_{Y_{N+1}|\mathcal{F}_N}(\tau) = (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^\top \nabla q_{N+1}(\boldsymbol{\gamma}_0, \tau) + o_p(N^{-1/2}),$$

where $\xrightarrow{L}$ represents the convergence in the distribution and $\boldsymbol{\Sigma}_2 = \sum_{k=1}^{K}\sum_{k'=1}^{K} \min\{\tau_k, \tau_{k'}\}(1 - \max\{\tau_k, \tau_{k'}\})$ $\mathrm{E}[\omega_t^2 \nabla q_t(\boldsymbol{\gamma}_0, \tau_k)\{\nabla q_t(\boldsymbol{\gamma}_0, \tau_{k'})\}^\top]$.

Theorem 2.2. Suppose that conditions **C1**–**C4** are satisfied. If $h = o(N^{-1/4})$, $h(N/\ln N)^{1/3} \to \infty$ and $N \to \infty$, then, we have

$$\sqrt{N}(\tilde{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}_0) \xrightarrow{L} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\right),$$

and

$$\tilde{Q}^*_{Y_{N+1}|\mathcal{F}_N}(\tau) - Q_{Y_{N+1}|\mathcal{F}_N}(\tau) = (\tilde{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}_0)^\top \nabla q_{N+1}(\boldsymbol{\gamma}_0, \tau) + o_p(N^{-1/2}),$$

where $\tilde{Q}^*_{Y_{N+1}|\mathcal{F}_N}(\tau) = \boldsymbol{X}_N^\top \tilde{\boldsymbol{\beta}}^* + \sigma(\boldsymbol{X}_N, \tilde{\boldsymbol{\alpha}}^*)\left(\tilde{\theta}_1^* + \tilde{\theta}_2^*\left[(\tau^{\tilde{\theta}_3^*} - 1)/\tilde{\theta}_3^* - \{(1-\tau)^{\tilde{\theta}_4^*} - 1\}/\tilde{\theta}_4^*\right]\right)$ and $\tilde{\boldsymbol{\gamma}}_b^* = (\tilde{\boldsymbol{\beta}}_b^{*\top}, \tilde{\boldsymbol{\alpha}}_b^{*\top}, \tilde{\theta}_{b,1}^*, \tilde{\theta}_{b,2}^*, \tilde{\theta}_{b,3}^*, \tilde{\theta}_{b,4}^*)^\top$.

Through the result of Theorems 2.1 and 2.2, the smoothing SWCQR (SSWCQR) estimator $\tilde{\boldsymbol{\gamma}}^*$ in (2.3) achieves optimal efficiency and its asymptotic covariance matrix, which is the same as that of SWCQR estimator $\hat{\boldsymbol{\gamma}}$ in (2.1).

3. RENEWABLE ESTIMATION METHOD FOR STREAMING DATASETS

3.1. Methodology

Now let us discuss how to use (2.3) to develop a renewable estimator for streaming data sets. Assume we have the streaming data sets $\{D_1, \ldots, D_b\}$ up to the bth batch, where D_j is the jth batch data set with a sample size of n_j. Then, the total sample size is $N_b = \sum_{j=1}^{b} n_j$. The data can exceed even a super computer's memory when the number

DOI: 10.1111/jtsa.12725

of the blocks b is large enough. For model (1.1), $D_j = \{\mathbf{X}_j, \mathbf{Y}_j\}$ is the jth batch data set, where $\mathbf{Y}_j = (Y_{1,j}, \ldots, Y_{n_j,j})^\top$ and $\mathbf{X}_j = (\mathbf{X}_{1,j}, \ldots, \mathbf{X}_{n_j,j})^\top$.

We begin with a simple scenario of two batches of data D_1 and D_2, where D_2 arrives after D_1. We want to update the initial SSWCQR $\tilde{\boldsymbol{\gamma}}_1$ (or $\tilde{\boldsymbol{\gamma}}_1^*$) by minimizing (2.3) to a renewable SSWCQR (RSSWCQR) $\tilde{\boldsymbol{\gamma}}$ without using any subject-level data but only some summary statistics from D_1. By (2.3) and (2.4), the SSWCQR $\tilde{\boldsymbol{\gamma}}_1$ satisfies,

$$\mathbf{U}(D_1; \tilde{\boldsymbol{\gamma}}_1; h_1) = \mathbf{0}, \tag{3.1}$$

where $\mathbf{U}(D_j; \boldsymbol{\gamma}; h) = \sum_{k=1}^K \sum_{t \in D_j} \omega_t \nabla q_t(\boldsymbol{\gamma}, \tau_k) \left\{K^*((q_t(\boldsymbol{\gamma}, \tau_k) - Y_t)/h) - \tau_k\right\}$.

We propose a new estimator $\tilde{\boldsymbol{\gamma}}_2$ for streaming data $\{D_1, D_2\}$ as a solution to the equation of the form

$$\mathbf{J}(D_1; \tilde{\boldsymbol{\gamma}}_1; h_1)(\tilde{\boldsymbol{\gamma}}_2 - \tilde{\boldsymbol{\gamma}}_1) + \mathbf{U}(D_2; \tilde{\boldsymbol{\gamma}}_2; h_2) = \mathbf{0}, \tag{3.2}$$

where $\mathbf{J}(D_j; \boldsymbol{\gamma}; h) = \partial \mathbf{U}(D_j; \boldsymbol{\gamma}; h)/\partial \boldsymbol{\gamma} = \sum_{k=1}^K \sum_{t \in D_j} \omega_t [\nabla^2 q_t(\boldsymbol{\gamma}, \tau_k)\{K^*((q_t(\boldsymbol{\gamma}, \tau_k) - Y_t)/h) - \tau_k\} + \nabla q_t(\boldsymbol{\gamma}, \tau_k) \left\{\nabla q_t(\boldsymbol{\gamma}, \tau_k)\right\}^\top K_h(q_t(\boldsymbol{\gamma}, \tau_k) - Y_t)]$. Equation (3.2) is according to

$$\begin{aligned}\mathbf{U}(D_1; \tilde{\boldsymbol{\gamma}}_2; h_1) &= \mathbf{U}(D_1; \tilde{\boldsymbol{\gamma}}_1; h_1) + \mathbf{J}(D_1; \tilde{\boldsymbol{\gamma}}_1; h_1)(\tilde{\boldsymbol{\gamma}}_2 - \tilde{\boldsymbol{\gamma}}_1) + O_p(n_1 \|\tilde{\boldsymbol{\gamma}}_2 - \tilde{\boldsymbol{\gamma}}_1\|_2^2) \\ &= \mathbf{J}(D_1; \tilde{\boldsymbol{\gamma}}_1; h_1)(\tilde{\boldsymbol{\gamma}}_2 - \tilde{\boldsymbol{\gamma}}_1) + O_p(n_1 \|\tilde{\boldsymbol{\gamma}}_2 - \tilde{\boldsymbol{\gamma}}_1\|_2^2),\end{aligned}$$

where the last equation is according to (3.1), $O_p(\cdot)$ means bounded with probability and the error term $O_p(n_1 \|\tilde{\boldsymbol{\gamma}}_2 - \tilde{\boldsymbol{\gamma}}_1\|_2^2)$ can be asymptotically ignored.

Through (3.2), the initial $\tilde{\boldsymbol{\gamma}}_1$ is renewed by $\tilde{\boldsymbol{\gamma}}_2$ only using the historical summary statistics, including sample variance matrix $\mathbf{J}(D_1; \tilde{\boldsymbol{\gamma}}_1; h_1)$ and estimate $\tilde{\boldsymbol{\gamma}}_1$, instead of the subject-level raw data D_1. Generalizing (3.2) to streaming data sets $\{D_1, \ldots, D_b\}$, a renewable estimator $\tilde{\boldsymbol{\gamma}}_b$ of $\boldsymbol{\gamma}_0$ is defined as a solution to the following incremental estimation equation:

$$\sum_{j=1}^{b-1} \mathbf{J}\left(D_j; \tilde{\boldsymbol{\gamma}}_j; h_j\right)\left(\tilde{\boldsymbol{\gamma}}_b - \tilde{\boldsymbol{\gamma}}_{b-1}\right) + \mathbf{U}\left(D_b; \tilde{\boldsymbol{\gamma}}_b; h_b\right) = \mathbf{0}. \tag{3.3}$$

Therefore, we can obtain the estimator $\tilde{Q}^b_{Y_t|\mathcal{F}_{t-1}}(\tau)$ of $Q_{Y_t|\mathcal{F}_{t-1}}(\tau)$ for any $\tau \in (0,1)$ and $\boldsymbol{X}_t$ in D_b based on $\tilde{\boldsymbol{\gamma}}_b = (\tilde{\boldsymbol{\beta}}_b^\top, \tilde{\boldsymbol{\alpha}}_b^\top, \tilde{\theta}_{b,1}, \tilde{\theta}_{b,2}, \tilde{\theta}_{b,3}, \tilde{\theta}_{b,4})^\top$ as

$$\tilde{Q}^b_{Y_t|\mathcal{F}_{t-1}}(\tau) = \boldsymbol{X}_t^\top \tilde{\boldsymbol{\beta}}_b + \sigma(\boldsymbol{X}_t, \tilde{\boldsymbol{\alpha}}_b)\left[\tilde{\theta}_{b,1} + \tilde{\theta}_{b,2}\left\{\frac{\tau^{\tilde{\theta}_{b,3}} - 1}{\tilde{\theta}_{b,3}} - \frac{(1-\tau)^{\tilde{\theta}_{b,4}} - 1}{\tilde{\theta}_{b,4}}\right\}\right]. \tag{3.4}$$

3.2. Large Sample Properties

The following theorem shows the asymptotic properties of the estimators $\tilde{\boldsymbol{\gamma}}_b$ in (3.3) and $\tilde{Q}^b_{Y_{N_b+1}|\mathcal{F}_{N_b}}(\tau)$ in (3.4).

Theorem 3.1. Suppose that conditions **C1–C4** are satisfied. If $h_j = o(N_j^{-1/4})$, $h_j(N_j/\ln N_j)^{1/3} \to \infty$ with $N_j = \sum_{i=1}^j n_i$ and $N_1 \to \infty$, then, we have

$$\sqrt{N_b}(\tilde{\boldsymbol{\gamma}}_b - \boldsymbol{\gamma}_0) \xrightarrow{L} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\right),$$

and

$$\tilde{Q}^{b}_{Y_{N_b+1}|\mathcal{F}_{N_b}}(\tau) - Q_{Y_{N_b+1}|\mathcal{F}_{N_b}}(\tau) = (\tilde{\boldsymbol{\gamma}}_b - \boldsymbol{\gamma}_0)^{\top}\nabla q_{N+1}(\boldsymbol{\gamma}_0, \tau) + o_p\left(N_b^{-1/2}\right).$$

Through the result of Theorems 2.2 and 3.1, it is interesting to notice that the renewable estimator $\tilde{\boldsymbol{\gamma}}_b$ achieves optimal efficiency and its asymptotic covariance matrix is the same as that of the SSWCQR estimator $\tilde{\boldsymbol{\gamma}}^*_b$ in (2.3), which is computed directly on all the samples. This implies that the proposed renewable estimator achieves the same asymptotic distribution as the SSWCQR estimator.

3.3. Algorithm

Numerically, it is quite straightforward to find $\tilde{\boldsymbol{\gamma}}_b$ from (3.3) using the Newton–Raphson method at the $(r+1)$th iteration:

$$\tilde{\boldsymbol{\gamma}}_b^{(r+1)} = \tilde{\boldsymbol{\gamma}}_b^{(r)} - \left\{\hat{\mathbf{J}}_{b-1} + \mathbf{J}(D_b; \tilde{\boldsymbol{\gamma}}_b^{(r)}; h_b)\right\}^{-1}\hat{\mathbf{U}}_b^{(r)}, \tag{3.5}$$

where $\hat{\mathbf{J}}_{b-1} = \sum_{j=1}^{b-1}\mathbf{J}(D_j; \tilde{\boldsymbol{\gamma}}_j; h_j)$ and $\hat{\mathbf{U}}_b^{(r)} = \hat{\mathbf{J}}_{b-1}(\tilde{\boldsymbol{\gamma}}_b^{(r)} - \tilde{\boldsymbol{\gamma}}_{b-1}) + \mathbf{U}(D_b; \tilde{\boldsymbol{\gamma}}_b^{(r)}; h_b)$. We summarize the general algorithm for the proposed RSSWCQR method by (3.5) as follows (Algorithm 1).

Algorithm 1. RSSWCQR estimation for streaming data sets

1: **Input:** streaming data sets $D_1, \ldots, D_b, \ldots$, the quantile levels $\tau_1, \ldots, \tau_K$,kernel function $K(\cdot)$ and bandwidths h_b with $b = 1, 2 \ldots$
2: **Initialize:** calculate $\tilde{\boldsymbol{\gamma}}_1$ by minimizing (2.3) with D_1 using the Newton–Raphson method as (3.5), and compute $\mathbf{J}(D_1; \tilde{\boldsymbol{\gamma}}_1; h_1)$;
3: **for:** $b = 2, 3, \ldots$ **do**
4: read in data set D_b;
5: select initial estimator $\tilde{\boldsymbol{\gamma}}_b^{(0)} = \tilde{\boldsymbol{\gamma}}_{b-1}$ anditerate (3.5) until convergence to obtain $\tilde{\boldsymbol{\gamma}}_b$;
6: update $\hat{\mathbf{J}}_b = \hat{\mathbf{J}}_{b-1} + \mathbf{J}(D_b; \tilde{\boldsymbol{\gamma}}_b; h_b)$;
7: save $\tilde{\boldsymbol{\gamma}}_b$ and $\hat{\mathbf{J}}_b$ and release data set D_b from the memory;
8: **end**
9: **Output:** $\tilde{\boldsymbol{\gamma}}_b$ for $b = 2, 3, \ldots$

Note that in step 5 in Algorithm 1 and (3.5), we only use the subject-level data of current data D_b and summary statistics $\hat{\mathbf{J}}_{b-1}$ and $\tilde{\boldsymbol{\gamma}}_{b-1}$ from historical data batches up to $b-1$ rather than subject-level raw data of $\{D_1, \ldots, D_{b-1}\}$ to obtain $\tilde{\boldsymbol{\gamma}}_b$. Thus, our proposed renewable method is indeed an online estimation procedure. Moreover, by step 7 in Algorithm 1, we only need to save $\tilde{\boldsymbol{\gamma}}_b$ and $\hat{\mathbf{J}}_b$ to obtain the next estimator, which are $p \times 1$ and $p \times p$, respectively. The scale of the data to be stored is $(p+1)p$ instead of $N_b p$, which is the sample size of the streaming data sets up to b batches. Because p is assumed to be a fixed number in this article, our method greatly reduces the amount of data storage.

4. NUMERICAL STUDIES

We first use Monte Carlo simulation studies to assess the finite sample performance of the proposed procedures and then demonstrate the application of the proposed methods with a real data analysis. All programs are written in R code. In all of the numerical experiments, we take $\omega_t = \left(1 + \sum_{j=1}^{p}|Y_{t-j}|^3\right)^{-1}$ for simplicity, which is also used in Zhu and Li (2022).

DOI: 10.1111/jtsa.12725

4.1. Simulation Example 1: SWCQR Estimation

We study the performance of the SWCQR estimation proposed in Section 2.1. We generate data from the following double-autoregressive model:

$$Y_t = 0.5Y_{t-1} + \varepsilon_t\sqrt{1 + 0.5Y_{t-1}^2}, \quad t = 1, \ldots, N. \tag{4.1}$$

Two error distributions of ε_t are considered: a standard normal distribution N(0, 1) and a t distribution with 3 degrees of freedom $t(3)$. The sample size is $N = 10{,}000$.

To evaluate the performance of the estimation method, we calculate the absolute error (AE) of β_0, α_0 and $Q_{\varepsilon_t}(\tau)$ as: $|\hat{\beta} - \beta_0|$, $|\hat{\alpha} - \alpha_0|$ and $|\hat{Q}_{\varepsilon_t}(\tau) - Q_{\varepsilon_t}(\tau)|$, where $\beta_0 = \alpha_0 = 0.5$ and $\hat{Q}_{\varepsilon_t}(\tau) = \hat{\theta}_1 + \hat{\theta}_2\left[(\tau^{\hat{\theta}_3} - 1)/\hat{\theta}_3 - \{(1-\tau)^{\hat{\theta}_4} - 1\}/\hat{\theta}_4\right]$. The true values of $Q_{\varepsilon_t}(\tau)$ for $\tau = 0.1, 0.5, 0.9$ are $-1.282, 0, 1.282$ and $-1.638, 0, 1.638$ for N(0, 1) and $t(3)$, respectively.

We first investigate the sensitivity of the proposed SWCQR method to the K in (2.1). The most commonly used K is 5, 9, and 19 (Zou and Yuan, 2008). The simulation results of AE are shown in Table I, which are based on 100 simulation replications. As can be seen from Table I that the AEs of β_0, α_0 and $Q_{\varepsilon_t}(\tau)$ are very close under different errors and Ks. Therefore, the SWCQR estimation is insensitive to K. Moreover, from Table I, we can see that all of the estimators are very close to the true values. In the following simulation, we will use $K = 5$ for convenience of calculation.

Next, we consider the performance of the SWCQR estimation for $Q_{Y_{N+1}|\mathcal{F}_N}(\tau)$ at $\tau \in \{0.1, 0.5, 0.9\}$. Moreover, we compare our method with QR method in Zhu and Li (2022). To evaluate the performances of the estimation methods, we calculate the mean absolute error (MAE $= N^{-1}\sum_{t=1}^{N}|\hat{Q}_{Y_t|\mathcal{F}_{t-1}}(\tau) - Q_{Y_t|\mathcal{F}_{t-1}}(\tau)|$). The simulation results of the MAE in Table II show that the performances of SWCQR are all better than those of QR under different quantile levels and errors.

Table I. The means and standard deviations (in parentheses) of AEs of β_0, α_0 and $Q_{\varepsilon_t}(\tau)$ under different Ks and errors for simulation example 1

Error	K	AE of β_0	AE of α_0	AE of $Q_{\varepsilon_t}(0.1)$	AE of $Q_{\varepsilon_t}(0.5)$	AE of $Q_{\varepsilon_t}(0.9)$
N(0, 1)	5	0.013 (0.008)	0.026 (0.020)	0.017 (0.011)	0.011 (0.009)	0.019 (0.014)
	9	0.013 (0.009)	0.027 (0.021)	0.017 (0.011)	0.011 (0.010)	0.019 (0.014)
	19	0.013 (0.009)	0.025 (0.019)	0.015 (0.012)	0.011 (0.009)	0.017 (0.012)
$t(3)$	5	0.017 (0.012)	0.041 (0.027)	0.034 (0.025)	0.014 (0.010)	0.042 (0.027)
	9	0.016 (0.012)	0.040 (0.026)	0.033 (0.024)	0.014 (0.010)	0.042 (0.028)
	19	0.016 (0.012)	0.039 (0.024)	0.030 (0.024)	0.013 (0.010)	0.040 (0.028)

Table II. The means and standard deviations (in parentheses) of MAEs under different methods, quantile levels and errors for simulation example 1

Error	Method	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
N(0, 1)	QR	0.538 (0.027)	0.034 (0.020)	0.052 (0.021)
	SWCQR	0.042 (0.023)	0.023 (0.013)	0.036 (0.016)
$t(3)$	QR	4.133 (2.323)	0.105 (0.068)	0.248 (0.148)
	SWCQR	0.189 (0.175)	0.081 (0.055)	0.169 (0.102)

DOI: 10.1111/jtsa.12725

4.2. Simulation Example 2: SSWCQR Estimation

We study the performance of the SSWCQR estimation proposed in Section 2.2. We generate data from the model (4.1).

We first investigate the sensitivity of the proposed SSWCQR method to bandwidth selection h in (2.3). From Theorem 2.2, we choose $h = CN^{-1/4}/\ln N$, which satisfy the condition for h. We vary the constant C from 0.01 to 100. To evaluate the performance of the estimation method, we calculate the root square error (RSE): $\text{RSE} = \sqrt{(\hat{\beta} - \beta_0)^2 + (\hat{\alpha} - \alpha_0)^2}$ and MAE of $Q_{Y_{N+1}|\mathcal{F}_N}(\tau)$ in Section 4.1.

The results are shown in Tables III and IV. As can be seen from Tables III and IV that the RSEs and MAEs of SWCQR and SSWCQR are very close under $C = 0.01$ and 0.1. Therefore, we take $h = 0.1N^{-1/4}/\ln N$ for SSWCQR method in the subsequent simulation study.

4.3. Simulation Example 3: RSSWCQR Estimation

We study the performance of the RSSWCQR estimation proposed in Section 3.

We generate data from the model (4.1), where only the $t(3)$ error is considered. Moreover, we fix the sample size of each batch to $n_j = 500$ for $j = 1, \ldots, b$ and vary the number of batches $b = 10, 50, 100, \ldots, 1000$. Simulation results are based on 100 simulation replications. By Theorem 3.1 and the analysis in simulation 4.2, we take $h_j = 0.1N_j^{-1/4}/\ln N_j$.

We compare our proposed RSSWCQR estimator with the following two competitors: (1) the SSWCQR estimator with full data; and (2) the average SSWCQR (ASSWCQR) estimator for the streaming data set, that is, estimate each streaming data separately and then take its average. To evaluate the performance of the three methods, we calculate the RSE in Section 4.2.

From Tables V and VI, we note that all the estimators are close to the true value because the RSEs are very small, and for any given number of batches b, the RSEs of the proposed estimator (RSSWCQR) are very close to those of SSWCQR and better than those of ASSWCQR.

Table III. The means and standard deviations (in parentheses) of RSEs and MAEs (×100) with different hs under N(0, 1) error for simulation example 2

Method	h	RSE	MAE ($\tau = 0.1$)	MAE ($\tau = 0.5$)	MAE ($\tau = 0.9$)
SWCQR	–	3.801 (2.392)	3.608 (1.756)	2.657 (1.586)	3.817 (1.625)
SSWCQR	0.01	3.793 (2.372)	3.605 (1.751)	2.659 (1.584)	3.816 (1.617)
	0.1	3.785 (2.393)	3.611 (1.767)	2.654 (1.589)	3.805 (1.619)
	1	3.904 (2.299)	4.066 (2.458)	2.707 (1.547)	4.420 (2.751)
	10	5.303 (5.775)	5.956 (4.143)	2.748 (1.765)	6.454 (4.379)
	100	33.263 (9.974)	34.580 (26.447)	14.552 (19.098)	28.354 (23.461)

Table IV. The means and standard deviations (in parentheses) of RSEs and MAEs (×100) with different hs under $t(3)$ error for simulation example 2

Method	h	RSE	MAE ($\tau = 0.1$)	MAE ($\tau = 0.5$)	MAE ($\tau = 0.9$)
SWCQR	–	5.026 (2.829)	19.148 (19.969)	9.285 (8.279)	19.217 (18.956)
SSWCQR	0.01	5.026 (2.828)	19.135 (19.942)	9.288 (8.295)	19.195 (18.821)
	0.1	5.021 (2.861)	19.552 (20.308)	9.290 (8.250)	19.437 (18.793)
	1	5.626 (4.964)	21.716 (24.196)	9.707 (8.804)	23.856 (27.371)
	10	6.499 (5.527)	28.491 (31.817)	11.750 (14.279)	30.227 (29.600)
	100	27.755 (13.860)	112.837 (112.735)	38.714 (68.155)	76.548 (62.371)

DOI: 10.1111/jtsa.12725

Table V. The means and standard deviations (in parentheses) of RSEs (×100) with different bs and methods under N(0, 1) error for simulation example 3

b	SSWCQR	RSSWCQR	ASSWCQR
10	5.054 (3.378)	6.507 (4.288)	5.502 (3.192)
50	2.280 (1.200)	2.653 (1.452)	2.503 (1.552)
100	1.278 (0.744)	1.528 (0.741)	1.830 (1.239)
200	1.180 (0.762)	1.317 (0.749)	1.643 (0.815)
300	0.985 (0.714)	1.091 (0.627)	1.539 (0.953)
400	0.826 (0.664)	0.915 (0.595)	1.493 (0.826)
500	0.721 (0.582)	0.793 (0.539)	1.474 (0.800)
600	0.641 (0.529)	0.745 (0.536)	1.448 (0.726)
700	0.630 (0.431)	0.711 (0.501)	1.474 (0.665)
800	0.591 (0.385)	0.661 (0.414)	1.444 (0.612)
900	0.540 (0.333)	0.646 (0.402)	1.391 (0.589)
1000	0.472 (0.314)	0.605 (0.397)	1.343 (0.531)

Table VI. The means and standard deviations (in parentheses) of RSEs (×100) with different bs and methods under $t(3)$ error for simulation example 3

b	SSWCQR	RSSWCQR	ASSWCQR
10	5.931 (4.083)	7.909 (5.294)	7.987 (6.081)
50	2.957 (2.045)	3.811 (2.424)	3.994 (2.445)
100	2.007 (1.353)	2.692 (1.818)	3.208 (1.933)
200	1.665 (1.029)	2.098 (1.202)	3.066 (1.616)
300	1.384 (0.861)	1.772 (1.232)	2.944 (1.507)
400	1.204 (0.770)	1.596 (1.030)	2.836 (1.333)
500	0.922 (0.646)	1.421 (0.841)	2.840 (1.123)
600	0.879 (0.640)	1.266 (0.819)	2.855 (1.139)
700	0.809 (0.599)	1.232 (0.787)	2.830 (1.044)
800	0.758 (0.525)	1.224 (0.774)	2.892 (0.956)
900	0.746 (0.466)	1.195 (0.726)	2.887 (0.954)
1000	0.682 (0.417)	1.100 (0.667)	2.966 (0.859)

Table VII. Summary statistics for S&P500 returns

Mean	Median	Std.dev.	Skewness	Kurtosis	Min	Max
0.033	0.059	1.138	−1.128	24.853	−22.918	10.964

4.4. Real Data Example: S&P500 Index Data

To illustrate the practical usefulness of application of our proposed methods, a daily data of S&P500 index between 1 January 1980 and 19 September 2022 with 10,772 observations in total. The data is downloaded from the website of Yahoo Finance (https://hk.finance.yahoo.com). The daily returns are computed as 100 times the difference of the log of the prices, that is, $Y_t = 100\ln(p_t/p_{t-1})$, where p_t is the daily price. Table VII collects the summary statistics of $\{Y_t\}$, where the sample skewness -1.128 indicates possible asymmetries in the volatility, and the sample kurtosis 24.853 implies heavy tail of $\{Y_t\}$. Figure 2 also gives the time series plot for S&P500.

DOI: 10.1111/jtsa.12725

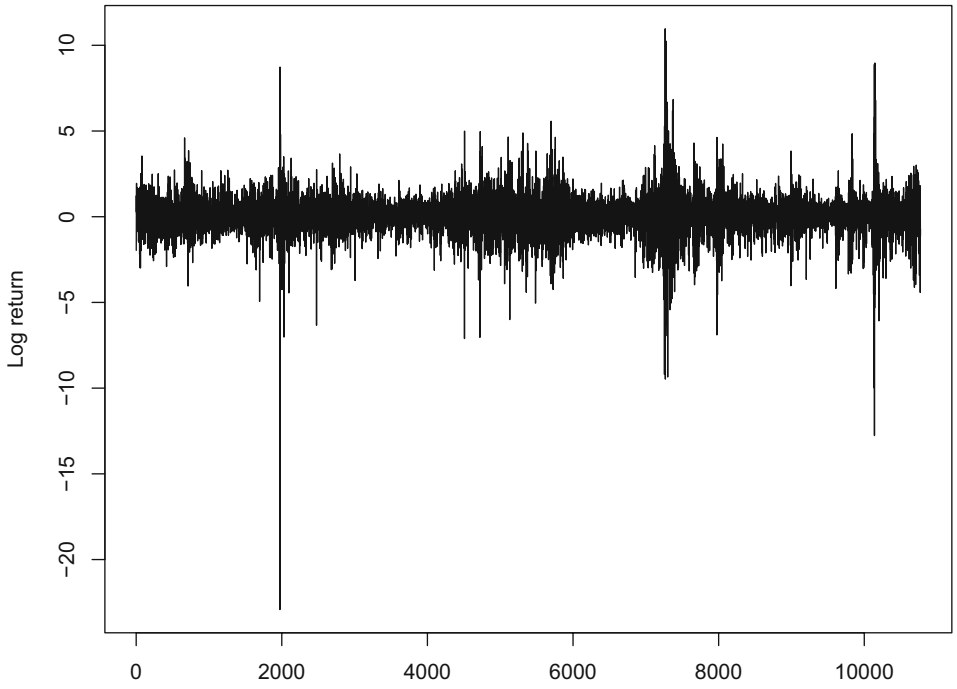


Figure 2. The time series plot of the daily return series: 1980–2022

We first consider the different order p for the following two double-autoregressive models ($p = 1, 2$):

$$Y_t = \beta Y_{t-1} + \varepsilon_t \sqrt{1 + \alpha Y_{t-1}^2}, \tag{4.2}$$

and

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_t \sqrt{1 + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2}. \tag{4.3}$$

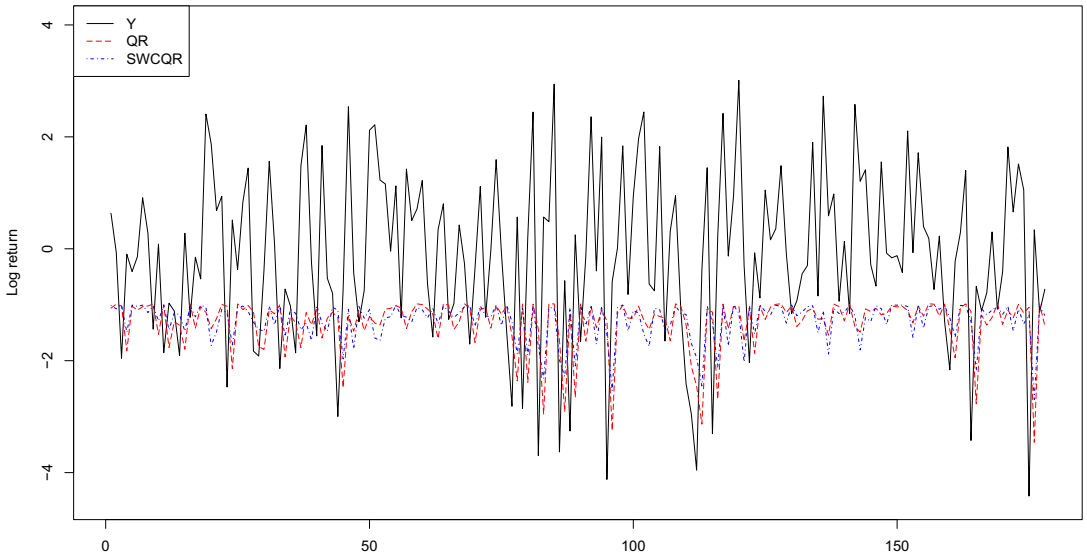
The mean absolute fitting error (MAFE=$N^{-1}\sum_{t=1}^{N} |Y_t - \boldsymbol{X}_t^{\top}\hat{\boldsymbol{\beta}}|$) is used to compare the fitting data of the model (under different p), where $N = 10772$ and $\hat{\boldsymbol{\beta}}$ is obtained by SWCQR. The results of MAFE under different models (4.2) and (4.3) are the same as 0.775. Therefore, in the following study, for convenience, we only consider model (4.2).

Since Value-at-Risk (VaR) is an important risk measure for financial assets, we use the model (4.2) to forest the conditional quantile of $\{Y_t\}$. The first 10,594 observations from 1980 to 2021 are used for model estimation and the remaining 178 observations are reserved for the out-of-sample evaluation. Figure 3 shows that the performances of QR and SWCQR are very close.

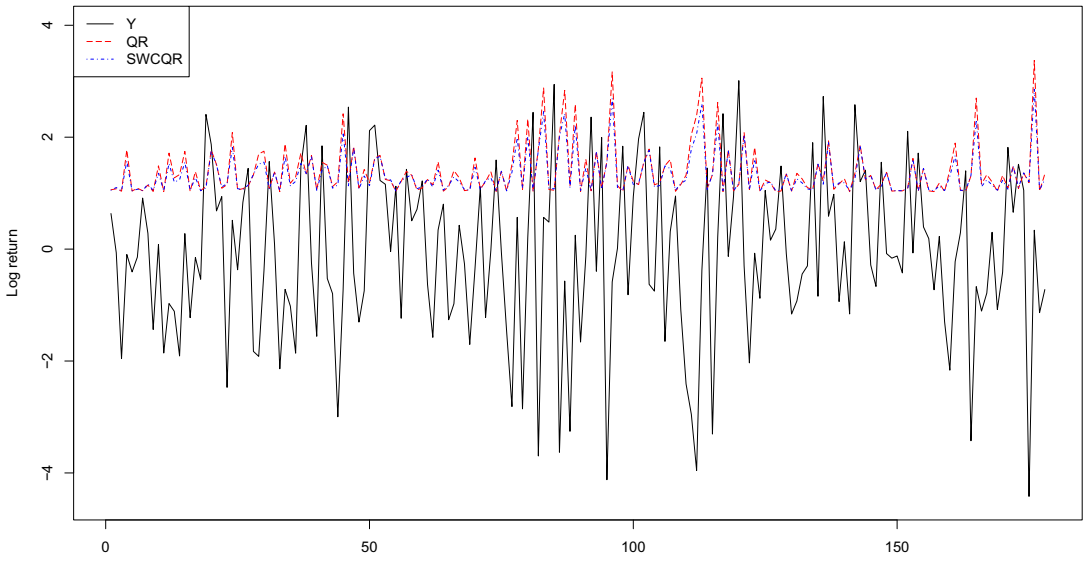
Next, we consider the proposed method RSSWCQR in Section 3. We consider $b = 42$ for the 42 years from 1980 to 2021, and the data of 1980 is regarded as the first streaming data. From Table VIII and Figure 4, we can see that the estimated coefficient of β and VaRs under $\tau = 0.1$ and 0.9 are all very close. Moreover, we calculate the relative absolute error (RAE) of SWCQR as

$$\text{RAE} = \frac{1}{N}\sum_{t=1}^{N} \frac{|\hat{Q}_{Y_t|\mathcal{F}_{t-1}}(\tau) - \hat{Q}^{SWCQR}_{Y_t|\mathcal{F}_{t-1}}(\tau)|}{|\hat{Q}^{SWCQR}_{Y_t|\mathcal{F}_{t-1}}(\tau)|} \times 100\%,$$

DOI: 10.1111/jtsa.12725



(a) VaR forecasts at the level of 10%



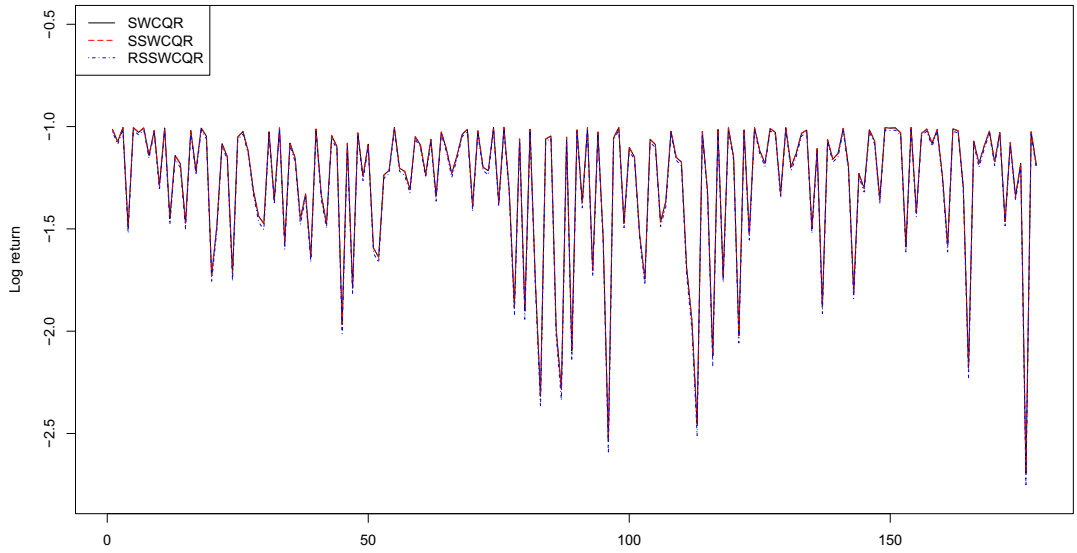
(b) VaR forecasts at the level of 90%

Figure 3. VaR forecasts at the level of 10% and 90% from 1 January 2022 to 19 September 2022 by QR and SWCQR methods. Y is the true value

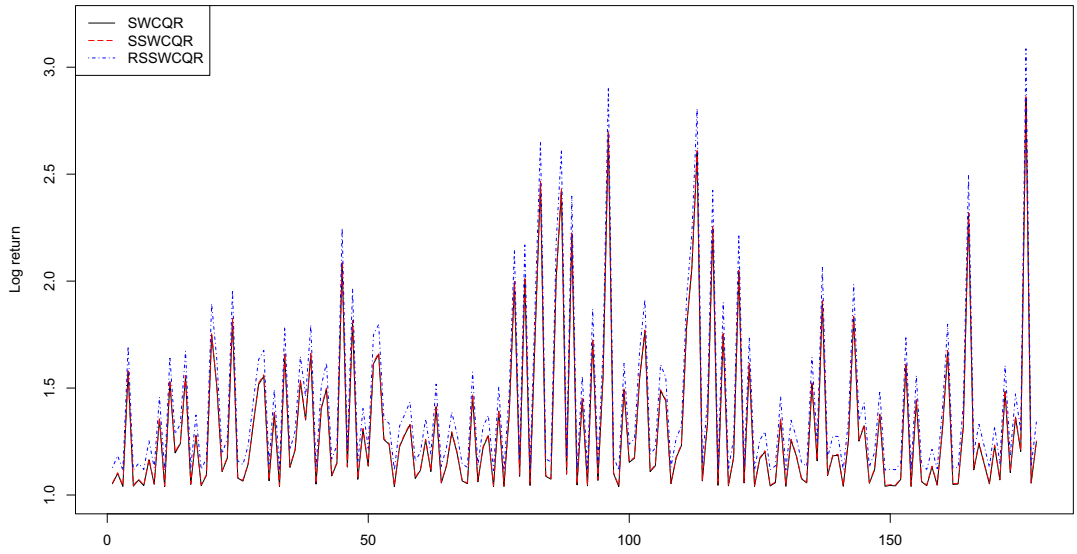
Table VIII. The estimated coefficient of β ($\times 10^{-3}$) by methods SWCQR, SSWCQR, RSSWCQR and ASSWCQR for model (4.2)

SWCQR	SSWCQR	RSSWCQR
−7.941	−8.235	−6.304

DOI: 10.1111/jtsa.12725



(a) VaR forecasts at the level of 10%



(b) VaR forecasts at the level of 90%

Figure 4. VaR forecasts at the level of 10% and 90% from 1 January 2022 to 19 September 2022 by SWCQR, SSWCQR, and RSSWCQR methods

where $\hat{Q}^{SWCQR}_{Y_t|\mathcal{F}_{t-1}}(\tau)$ is obtained by SWCQR and the data are from 2022. RAEs at $\tau = 0.1$ and 0.9 of SSWCQR are 0.070% and 0.358%, respectively. RAEs at $\tau = 0.1$ and 0.9 of RSSWCQR are 1.450% and 7.647%, respectively. The results of RAEs also show SSWCQR and RSSWCQR are close to SWCQR.

5. CONCLUSION

In this article, we considered a renewable non-crossing QR for the DAR model with streaming datasets by a parameterized method. In this way, the proposed quantile DAR model allows us to study the whole conditional

DOI: 10.1111/jtsa.12725

distribution of financial returns and then obtain the corresponding multi-step ahead conditional predictive distributions. Moreover, the method requires only the availability of the current data batch in the data stream and sufficient statistics on the historical data (the latest estimator and the cumulative Hessian matrix) in each stage of the analysis. The scale of the data to be stored is $(p + 1)p$ instead of $N_b p$, which is the sample size of streaming data sets up to b batches. Because p is assumed to be a fixed number in this article, our method greatly reduces the amount of data storage. Theoretically, the proposed estimators achieve optimal efficiency, and their asymptotic covariance matrices are the same as those of the estimators with full data. The algorithm of the proposed renewable non-crossing QR estimation is fast and scalable due to a convolution-type smoothing approach of the objective function.

From the numerical studies in Section 4, we can see that the parameterization of quantile functions by GLD works very well, and our proposed renewable method is very close to the estimator directly using all data.

ACKNOWLEDGMENTS

This research is supported by the National Social Science Foundation of China (Series number: 21BTJ040).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Yahoo Finance at https://hk.finance.yahoo.com.

REFERENCES

Cai Y. 2016. A general quantile function model for economic and financial time series. *Econometric Reviews* **35**:1173–1193. https://doi.org/10.1080/07474938.2014.976528.

Cai Y, Li G. 2019. A quantile function approach to the distribution of financial returns following TGARCH models. *Statistical Modelling* **21**:1471082X1987637. https://doi.org/10.1177/1471082X19876371.

Cai Y, Montes-Rojas G, Olmo J. 2013. Quantile double AR time series models for financial returns. *Journal of Forecasting* **32**:551–560. https://doi.org/10.1002/for.2261.

Chen X, Liu W, Zhang Y. 2019. Quantile regression under memory constraint. *Annals of Statistics* **47**:3244–3273. https://doi.org/10.1214/18-AOS1777.

Dedduwakumara DS, Prendergast LA, Staudte RG. 2021. An efficient estimator of the parameters of the generalized lambda distribution. *Journal of Statistical Computation and Simulation* **91**:197–215. https://doi.org/10.1080/00949655.2020.1808979.

Fernandes M, Guerre E, Horta E. 2021. Smoothing quantile regressions. *Journal of Business & Economic Statistics* **39**:338–357. https://doi.org/10.1080/07350015.2019.1660177.

Freimer M, Kollia G, Mudholkar G, Lin C. 1988. A study of the generalized tukey lambda family. *Communications in Statistics - Theory and Methods* **17**:3547–3567. https://doi.org/10.1080/03610928808829820.

Frumento P, Bottai M. 2016. Parametric modeling of quantile regression coefficient functions. *Biometrics* **72**:74–84. https://doi.org/10.1111/biom.12410.

Frumento P, Bottai M, Fernández-Val I. 2021. Parametric modeling of quantile regression coefficient functions with longitudinal data. *Journal of the American Statistical Association* **116**:783–797. https://doi.org/10.1080/01621459.2021.1892702.

Han R, Luo L, Lin Y, Huang J. 2021. Online debiased lasso. arXiv:2106.05925v1.

Jiang R, Yu K. 2022. Renewable quantile regression for streaming data sets. *Neurocomputing* **508**:208–224.

Jiang R, Yu K. 2023. No-crossing single-index quantile regression curve estimation. *Journal of Business & Economic Statistics* **41**:309–320. https://doi.org/10.1080/07350015.2021.2013245.

Karian Z, Dudewicz E. 2000. *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods* CRC Press, Boca Raton, FL.

Knight K. 1998. Limiting distributions for L1 regression estimators under general conditions. *Annals of Statistics* **26**:755–770.

Koenker R, Bassett G. 1978. Regression quantile. *Econometrica* **46**:33–50. https://doi.org/10.2307/1913643.

Li D, Ling S, Zakoı¨an J-M. 2015. Asymptotic inference in multiple-threshold double autoregressive models. *Journal of Econometrics* **189**:415–427. https://doi.org/10.1016/j.jeconom.2015.03.033.

Ling S. 2004. Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society: Series B* **66**:63–78. https://doi.org/10.1111/j.1467-9868.2004.00432.x.
Ling S. 2005. Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B* **67**:381–393. https://doi.org/10.1111/j.1467-9868.2005.00507.x.
Ling S. 2007. A double AR(*p*) model: structure and estimation. *Statistica Sinica* **17**:161–175.
Luo L, Song P. 2020. Renewable estimation and incremental inference in generalized linear models with streaming data sets. *Journal of the Royal Statistical Society: Series B* **82**:69–97. https://doi.org/10.1111/rssb.12352.
Luo L, Zhou L, Song P. 2022. Real-time regression analysis of streaming clustered data with possible abnormal data batches. *Journal of the American Statistical Association*. **118**:2029–2044. https://doi.org/10.1080/01621459.2022.2026778.
Pollard D. 1991. Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**:186–199.
Quan M, Lin Z. 2022. Optimal one-pass nonparametric estimation under memory constraint. *Journal of the American Statistical Association*. **0**:1–12. https://doi.org/10.1080/01621459.2022.2115374.
Schifano E, Wu J, Wang C, Yan J, Chen M. 2016. Online updating of statistical inference in the big data setting. *Technometrics* **58**:393–403. https://doi.org/10.1080/00401706.2016.1142900.
Wang K, Wang H, Li S. 2022. Renewable quantile regression for streaming datasets. *Knowledge-Based Systems* **235**:107675. https://doi.org/10.1016/j.knosys.2021.107675.
Xu Z, Zhao Z. 2021. Efficient estimation for models with nonlinear heteroscedasticity. *Journal of Business & Economic Statistics*. **40**:1498–1508. https://doi.org/10.1080/07350015.2021.1933991.
Yang Y, Yao F. 2022. Online estimation for functional data. *Journal of the American Statistical Association*. **118**:1630–164. https://doi.org/10.1080/01621459.2021.2002158.
Zhang Y, Si Y, Li G, Tsai C-L. 2021. Quantile index regression. arXiv:2111.03223.
Zhu Q, Li G. 2022. Quantile double autoregression. *Econometric Theory* **38**:793–839. https://doi.org/10.1017/S026646662100030X.
Zou H, Yuan M. 2008. Composite quantile regression and the oracle model selection theory. *Annals of Statistics* **36**:1108–1126. https://doi.org/10.1214/07-AOS507.

APPENDIX A: PROOF OF MAIN RESULTS

Proof of Theorem 2.1. Let $\boldsymbol{\mu} = \sqrt{N}\left(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\right)$ and $v_{tk} = q_t\left(\boldsymbol{\gamma}, \tau_k\right) - q_t\left(\boldsymbol{\gamma}_0, \tau_k\right)$. By Taylor expansion, we have $v_{tk} = \boldsymbol{\mu}^{\top}\nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right)/\sqrt{N} + \boldsymbol{\mu}^{\top}\nabla^2 q_t\left(\boldsymbol{\gamma}^*, \tau_k\right)\boldsymbol{\mu}/(2N)$, where $\boldsymbol{\gamma}^*$ is between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_0$. Then $\hat{\boldsymbol{\mu}} = \sqrt{N}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right)$ is also the minimizer of the following criterion:

$$L_N(\boldsymbol{\mu}) = \sum_{k=1}^{K}\sum_{t=p+1}^{N}\omega_t\left\{\rho_{\tau_k}\left(e_{tk} - \boldsymbol{\mu}^{\top}\nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right)/\sqrt{N} - \boldsymbol{\mu}^{\top}\nabla^2 q_t\left(\boldsymbol{\gamma}^*, \tau_k\right)\boldsymbol{\mu}/(2N)\right) - \rho_{\tau_k}\left(e_{tk}\right)\right\},$$

where $e_{tk} = Y_t - q_t\left(\boldsymbol{\gamma}_0, \tau_k\right)$. By the identity (Knight, 1998),

$$\rho_\tau(x-y) - \rho_\tau(x) = y\left\{\mathrm{I}(x<0) - \tau\right\} + y\int_0^1\left\{\mathrm{I}(x \le ys) - \mathrm{I}(x \le 0)\right\}\mathrm{d}s,$$

$L_N(\boldsymbol{\mu})$ can be rewritten as

$$L_N(\boldsymbol{\mu}) = \boldsymbol{\mu}^{\top}\mathbf{U}_N + \mathrm{V}_N(\boldsymbol{\mu}) + \mathrm{R}_1 + \mathrm{R}_2 + \mathrm{R}_3, \tag{A1}$$

where

$$\mathbf{U}_N = \frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{t=p+1}^{N}\omega_t\nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right)\left\{\mathrm{I}\left(e_{tk} < 0\right) - \tau\right\},$$

DOI: 10.1111/jtsa.12725

$$
\begin{aligned}
\mathrm{V}_N(\boldsymbol{\mu}) &= \boldsymbol{\mu}^\top \frac{1}{\sqrt{N}} \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right) \int_0^1 \left\{ \mathrm{I}\left(e_{tk} \le \boldsymbol{\mu}^\top \nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right) s\right) - \mathrm{I}\left(e_{tk} \le 0\right) \right\} \mathrm{d}s, \\
\mathrm{R}_1 &= \boldsymbol{\mu}^\top \frac{1}{2N} \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \nabla^2 q_t\left(\boldsymbol{\gamma}^*, \tau_k\right) \left\{ \mathrm{I}\left(e_{tk} < 0\right) - \tau \right\} \boldsymbol{\mu}, \\
\mathrm{R}_2 &= \boldsymbol{\mu}^\top \frac{1}{\sqrt{N}} \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right) \int_0^1 \left\{ \mathrm{I}\left(e_{tk} \le \nu_{tk} s\right) - \mathrm{I}\left(e_{tk} \le \boldsymbol{\mu}^\top \nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right) s\right) \right\} ds, \\
\mathrm{R}_3 &= \boldsymbol{\mu}^\top \frac{1}{2N} \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \nabla^2 q_t\left(\boldsymbol{\gamma}^*, \tau_k\right) \int_0^1 \left\{ \mathrm{I}\left(e_{tk} \le \nu_{tk} s\right) - \mathrm{I}\left(e_{tk} \le 0\right) \right\} \mathrm{d}s \boldsymbol{\mu}.
\end{aligned}
$$

Note that,

$$
\begin{aligned}
\mathrm{E}\left\{\mathrm{V}_N(\boldsymbol{\mu}) \,|\mathcal{F}_{t-1}\right\} &= \boldsymbol{\mu}^\top \frac{1}{\sqrt{N}} \sum_{k=1}^{K} \sum_{t=p+1}^{N} \omega_t \nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right) \int_0^1 \left\{ \mathrm{F}_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0, \tau_k\right) + \boldsymbol{\mu}^\top \nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right) s\right) \right. \\
&\quad \left. - \mathrm{F}_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0, \tau_k\right)\right) \right\} \mathrm{d}s \\
&= \boldsymbol{\mu}^\top \frac{1}{2N} \sum_{k=1}^{K} \sum_{t=p+1}^{N} f_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0, \tau_k\right)\right) \omega_t \nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right) \nabla q_t\left(\boldsymbol{\gamma}_0, \tau_k\right)^\top \boldsymbol{\mu} + o_p\left(\|\boldsymbol{\mu}\|_2^2\right), \\
&= \frac{1}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\mu} + o_p\left(\|\boldsymbol{\mu}\|_2^2\right).
\end{aligned}
$$

By Lemmas 1 and 2 in Zhu and Li (2022), we can obtain

$$
\begin{aligned}
\mathrm{V}_N(\boldsymbol{\mu}) &= \mathrm{E}\left\{\mathrm{V}_N(\boldsymbol{\mu}) \,|\mathcal{F}_{t-1}\right\} + \left[\mathrm{V}_N(\boldsymbol{\mu}) - \mathrm{E}\left\{\mathrm{V}_N(\boldsymbol{\mu}) \,|\mathcal{F}_{t-1}\right\}\right] \\
&= \frac{1}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\mu} + o_p\left(\|\boldsymbol{\mu}\|_2 + \|\boldsymbol{\mu}\|_2^2\right), \\
\mathrm{R}_1 + \mathrm{R}_2 + \mathrm{R}_3 &= o_p\left(\|\boldsymbol{\mu}\|_2 + \|\boldsymbol{\mu}\|_2^2\right). \qquad \text{(A2)}
\end{aligned}
$$

Thus, by (A1) and (A2), we have

$$
L_N(\boldsymbol{\mu}) = \boldsymbol{\mu}^\top \mathbf{U}_N + \frac{1}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\mu} + o_p\left(\|\boldsymbol{\mu}\|_2 + \|\boldsymbol{\mu}\|_2^2\right).
$$

It follows by the convexity lemma (Pollard, 1991) that the quadratic approximation to $L_N(\boldsymbol{\mu})$ holds uniformly for $\boldsymbol{\mu}$ in any compact set. Thus, it follows that

$$
\sqrt{N}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right) = \hat{\boldsymbol{\mu}} = -\boldsymbol{\Sigma}_1^{-1} \mathbf{U}_N + o_p(1) \xrightarrow{L} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1}\right). \qquad \text{(A3)}
$$

By (A3), we can obtain

$$
\begin{aligned}
\hat{Q}_{Y_{N+1}|\mathcal{F}_N}(\tau) - Q_{Y_{N+1}|\mathcal{F}_N}(\tau) &= q_{N+1}(\hat{\boldsymbol{\gamma}}, \tau) - q_{N+1}\left(\boldsymbol{\gamma}_0, \tau\right) \\
&= \left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right)^\top \nabla q_{N+1}\left(\boldsymbol{\gamma}_0, \tau\right) + o_p\left(N^{-1/2}\right). \qquad \text{(A4)}
\end{aligned}
$$

This completes the proof by (A3) and (A4). ■

Proof of Theorem 2.2. Denote $\tilde{\boldsymbol{\mu}}^* = \sqrt{N}\left(\tilde{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}_0\right)$, which is also the minimizer of the following criterion:

$$\begin{aligned}\tilde{L}_N(\boldsymbol{\mu}) = &\sum_{k=1}^{K}\sum_{t=p+1}^{N}\omega_t\int_{-\infty}^{+\infty}\rho_{\tau_k}(s)\,K_h\Big(s - e_{tk} + \boldsymbol{\mu}^\top\nabla q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)/\sqrt{N}\\ &+ \boldsymbol{\mu}^\top\nabla^2 q_t\left(\boldsymbol{\gamma}^*,\tau_k\right)\boldsymbol{\mu}/(2N)\Big)\mathrm{d}s - \sum_{k=1}^{K}\sum_{t=p+1}^{N}\omega_t\int_{-\infty}^{+\infty}\rho_{\tau_k}(s)\,K_h\left(s - e_{tk}\right)\mathrm{d}s\\ = &\boldsymbol{\mu}^\top\frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{t=p+1}^{N}\omega_t\nabla q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\left\{K^*\left(-e_{tk}/h\right) - \tau_k\right\}\\ &+ \boldsymbol{\mu}^\top\frac{1}{2N}\sum_{k=1}^{K}\sum_{t=p+1}^{N}\omega_t\nabla q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\left\{\nabla q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\right\}^\top K_h\left(-e_{tk}\right)\boldsymbol{\mu} + o_p\left(\|\boldsymbol{\mu}\|_2^2\right) + O_p\left(\|\boldsymbol{\mu}\|_2^3\right),\end{aligned} \tag{A5}$$

where the last equation in (A5) is similar to the proof of Theorem 2.1. Moreover, by conditions **C1** and **C4**, we can obtain

$$\begin{aligned}&\mathrm{E}\left\{K_h\left(-e_{tk}\right)|\mathcal{F}_{t-1}\right\} - f_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\right)\\ &= \mathrm{E}\left\{K_h\left(q_t\left(\boldsymbol{\gamma}_0,\tau_k\right) - Y_t\right)|\mathcal{F}_{t-1}\right\} - f_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\right)\\ &= \int_{-\infty}^{+\infty}K_h\left(q_t\left(\boldsymbol{\gamma}_0,\tau_k\right) - y\right)f\left(y|\mathcal{F}_{t-1}\right)dy - f_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\right)\\ &= \int_{-\infty}^{+\infty}K(s)\left\{f_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0,\tau_k\right) + hs\right) - f_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\right)\right\}\mathrm{d}s\\ &= O\left(h^2\right).\end{aligned} \tag{A6}$$

By (A5) and (A6), we have

$$\begin{aligned}\tilde{L}_N(\boldsymbol{\mu}) = &\boldsymbol{\mu}^\top\frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{t=p+1}^{N}\omega_t\nabla q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\left\{K^*\left(-e_{tk}/h\right) - \tau_k\right\}\\ &+ \frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{\Sigma}_1\boldsymbol{\mu} + o_p\left(\|\boldsymbol{\mu}\|_2^2\right) + O_p\left(\|\boldsymbol{\mu}\|_2^3\right).\end{aligned}$$

Thus, by the convexity lemma, we have

$$\sqrt{N}\left(\tilde{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}_0\right) = \tilde{\boldsymbol{\mu}}^* = -\boldsymbol{\Sigma}_1^{-1}\frac{1}{\sqrt{N}}\sum_{k=1}^{K}\sum_{t=p+1}^{N}\omega_t\nabla q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\left\{K^*\left(-e_{tk}/h\right) - \tau_k\right\} + o_p(1). \tag{A7}$$

Next, we consider that

$$\begin{aligned}\mathrm{E}\left\{K^*\left(-e_{tk}/h\right)|\boldsymbol{X}_t\right\} &= \int_{-\infty}^{q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)}\mathrm{E}\left\{K_h\left(s - Y_t\right)|\mathcal{F}_{t-1}\right\}ds\\ &= \int_{-\infty}^{q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)}f_{Y_t|\mathcal{F}_{t-1}}(s)\,\mathrm{d}s + O\left(h^2\right)\\ &= \tau_k + O\left(h^2\right),\end{aligned} \tag{A8}$$

DOI: 10.1111/jtsa.12725

and

$$\mathrm{E}\left[\left\{K^*\left(-e_{tk}/h\right)\right\}^2|\mathcal{F}_{t-1}\right] = \tau_k - 2hf_{Y_t|\mathcal{F}_{t-1}}\left(q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\right)\int_0^\infty K^*(s)\left\{1-K^*(s)\right\}\mathrm{d}s + O\left(h^2\right)$$
$$= \tau_k + O(h). \tag{A9}$$

Then, by (A7)–(A9) and condition $h = o\left(N^{-1/4}\right)$, we have

$$\sqrt{N}\left(\tilde{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}_0\right) \xrightarrow{L} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\right). \tag{A10}$$

This completes the proof by (A4) and (A10). ■

Proof of Theorem 3.1. Define a function

$$\mathbf{G}_b(\boldsymbol{\gamma}) = \sum_{j=1}^{b-1}\mathbf{J}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right)\left(\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}_{b-1}\right) + \mathbf{U}\left(D_b;\boldsymbol{\gamma};h_b\right). \tag{A11}$$

Thus, we have

$$\mathbf{G}_b\left(\tilde{\boldsymbol{\gamma}}_b\right) - \mathbf{G}_b\left(\boldsymbol{\gamma}_0\right) = \sum_{j=1}^{b-1}\mathbf{J}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right)\left(\tilde{\boldsymbol{\gamma}}_b-\boldsymbol{\gamma}_0\right) + \mathbf{U}\left(D_b;\tilde{\boldsymbol{\gamma}}_b;h_b\right) - \mathbf{U}\left(D_b;\boldsymbol{\gamma}_0;h_b\right)$$
$$= \left\{\sum_{j=1}^{b-1}\mathbf{J}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right) + \mathbf{J}\left(D_b;\boldsymbol{\gamma}_0;h_b\right)\right\}\left(\tilde{\boldsymbol{\gamma}}_b-\boldsymbol{\gamma}_0\right) + O_p\left(n_b\|\tilde{\boldsymbol{\gamma}}_b-\boldsymbol{\gamma}_0\|_2^2\right). \tag{A12}$$

According to (3.3), the renewable estimator $\tilde{\boldsymbol{\gamma}}_b$ satisfies $\mathbf{G}_b\left(\tilde{\boldsymbol{\gamma}}_b\right) = \mathbf{0}$. Thus, from (A11) and (A12), we can obtain

$$\left\{\sum_{j=1}^{b-1}\mathbf{J}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right) + \mathbf{J}\left(D_b;\boldsymbol{\gamma}_0;h_b\right)\right\}\left(\tilde{\boldsymbol{\gamma}}_b-\boldsymbol{\gamma}_0\right) + \sum_{j=1}^{b-1}\mathbf{J}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right)\left(\boldsymbol{\gamma}_0-\tilde{\boldsymbol{\gamma}}_{b-1}\right)$$
$$+ \mathbf{U}\left(D_b;\boldsymbol{\gamma}_0;h_b\right) + O_p\left(n_b\|\tilde{\boldsymbol{\gamma}}_b-\boldsymbol{\gamma}_0\|_2^2\right) = \mathbf{0}. \tag{A13}$$

By $\mathbf{U}\left(D_1;\tilde{\gamma}_1;h_1\right) = \mathbf{0}$, we have

$$\mathbf{U}\left(D_1;\boldsymbol{\gamma}_0;h_1\right) = \mathbf{U}\left(D_1;\tilde{\boldsymbol{\gamma}}_1;h_1\right) + \mathbf{J}\left(D_1;\tilde{\boldsymbol{\gamma}}_1;h_1\right)\left(\boldsymbol{\gamma}_0-\tilde{\boldsymbol{\gamma}}_1\right) + O_p\left(n_1\|\tilde{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_0\|_2^2\right),$$
$$= \mathbf{J}\left(D_1;\tilde{\boldsymbol{\gamma}}_1;h_1\right)\left(\boldsymbol{\gamma}_0-\tilde{\boldsymbol{\gamma}}_1\right) + O_p\left(n_1\|\tilde{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_0\|_2^2\right). \tag{A14}$$

By (3.2), we can obtain

$$\mathbf{U}\left(D_2;\boldsymbol{\gamma}_0;h_2\right) = \mathbf{U}\left(D_2;\tilde{\boldsymbol{\gamma}}_2;h_2\right) + \mathbf{J}\left(D_2;\tilde{\boldsymbol{\gamma}}_2;h_2\right)\left(\boldsymbol{\gamma}_0-\tilde{\boldsymbol{\gamma}}_2\right) + O_p\left(n_2\|\tilde{\boldsymbol{\gamma}}_2-\boldsymbol{\gamma}_0\|_2^2\right)$$
$$= -\mathbf{J}\left(D_1;\tilde{\boldsymbol{\gamma}}_1;h_1\right)\left(\tilde{\boldsymbol{\gamma}}_2-\tilde{\boldsymbol{\gamma}}_1\right) + \mathbf{J}\left(D_2;\tilde{\boldsymbol{\gamma}}_2;h_2\right)\left(\boldsymbol{\gamma}_0-\tilde{\boldsymbol{\gamma}}_2\right) + O_p\left(n_2\|\tilde{\boldsymbol{\gamma}}_2-\boldsymbol{\gamma}_0\|_2^2\right) \tag{A15}$$

Thus, combining (A14) and (A15),

$$\mathbf{U}\left(D_1;\boldsymbol{\gamma}_0;h_1\right) + \mathbf{U}\left(D_2;\boldsymbol{\gamma}_0;h_2\right) = \sum_{j=1}^{2}\mathbf{U}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right)\left(\boldsymbol{\gamma}_0-\tilde{\boldsymbol{\gamma}}_2\right) + \sum_{j=1}^{2}O_p\left(n_j\|\tilde{\boldsymbol{\gamma}}_j-\boldsymbol{\gamma}_0\|_2^2\right). \tag{A16}$$

Similarly to (A16), at the $(b-1)$th data batch, it is easy to show that

$$\sum_{j=1}^{b-1}\mathbf{U}\left(D_j;\boldsymbol{\gamma}_0;h_j\right)=\sum_{j=1}^{b-1}\mathbf{J}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right)\left(\boldsymbol{\gamma}_0-\tilde{\boldsymbol{\gamma}}_{b-1}\right)+\sum_{j=1}^{b-1}O_p\left(n_j\|\tilde{\boldsymbol{\gamma}}_j-\boldsymbol{\gamma}_0\|_2^2\right). \tag{A17}$$

Plugging (A17) into (A13), we get

$$\begin{aligned}&\left\{\sum_{j=1}^{b-1}\mathbf{J}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right)+\mathbf{J}\left(D_b;\boldsymbol{\gamma}_0;h_b\right)\right\}\left(\tilde{\boldsymbol{\gamma}}_b-\boldsymbol{\gamma}_0\right)+\sum_{j=1}^{b}\mathbf{U}\left(D_j;\boldsymbol{\gamma}_0;h_j\right)\\&+\sum_{j=1}^{b}O_p\left(n_j\|\tilde{\boldsymbol{\gamma}}_j-\boldsymbol{\gamma}_0\|_2^2\right)=\mathbf{0}.\end{aligned} \tag{A18}$$

By Theorem 2.2 and $N_1\to\infty$, $\tilde{\boldsymbol{\gamma}}_1$ is $\sqrt{N_1}$-consistent. If $\{\tilde{\boldsymbol{\gamma}}_j\}_{j=1}^{b-1}$ are $\sqrt{N_j}$-consistent, we can proof that $\tilde{\boldsymbol{\gamma}}_b$ is $\sqrt{N_b}$-consistent. Thus, by the proof of Theorem 2.2 and conditions $h_j\to 0$ and $h_jN_j\to\infty$, we have

$$\begin{aligned}&\frac{1}{N_b}\left\{\sum_{j=1}^{b-1}\mathbf{J}\left(D_j;\tilde{\boldsymbol{\gamma}}_j;h_j\right)+\mathbf{J}\left(D_b;\boldsymbol{\gamma}_0;h_b\right)\right\}\\&=\boldsymbol{\Sigma}_1+\frac{1}{N_b}O_p\left(\sum_{j=1}^{b}n_jh_j^2+\sqrt{\sum_{j=1}^{b}n_j/h_j}\right)\\&=\boldsymbol{\Sigma}_1+o_p\left(1\right).\end{aligned} \tag{A19}$$

By (A8), (A9), the Lemma 1 in Fernandes *et al.* (2021) and condition $h_j=o\left(N_j^{-1/4}\right)$, we have

$$\begin{aligned}\frac{1}{N_b}\sum_{j=1}^{b}\mathbf{U}\left(D_j;\boldsymbol{\gamma}_0;h_j\right)&=\mathbf{T}_b+\frac{1}{N_b}O_p\left(\sum_{j=1}^{b}n_jh_j^2+\sqrt{\sum_{j=1}^{b}n_jh_j}\right)\\&=\mathbf{T}_b+o_p\left(N_b^{-1/2}\right),\end{aligned} \tag{A20}$$

where $\mathbf{T}_b=N_b^{-1}\sum_{j=1}^{b}\sum_{i\in D_j}\sum_{k=1}^{K}\omega_t\nabla q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)\left\{I\left(Y_i-q_t\left(\boldsymbol{\gamma}_0,\tau_k\right)<0\right)-\tau_k\right\}$. The last equation in (A20) is according to $\sum_{j=1}^{b}n_j/N_j\le 1+\log\left(N_b/N_1\right)$ and $\sum_{j=1}^{b}n_j/\sqrt{N_j}\le 2\sqrt{N_b}$, which is the Lemma 3 in Han *et al.* (2021). Plugging (A19) and (A20) into (A18), we can obtain

$$\left(\boldsymbol{\Sigma}_1+o_p\left(1\right)\right)\left(\tilde{\boldsymbol{\gamma}}_b-\boldsymbol{\gamma}_0\right)+\mathbf{T}_b+\frac{n_b}{N_b}O_p\left(\|\tilde{\boldsymbol{\gamma}}_b-\boldsymbol{\gamma}_0\|_2^2\right)+o_p\left(N_b^{-1/2}\right)=\mathbf{0}. \tag{A21}$$

Then, we can proof that $\tilde{\boldsymbol{\gamma}}_b$ is $\sqrt{N_b}$-consistent. Finally, by (A21), Theorem 3.1 can be proved in a similar way to Theorem 2.2. ■

DOI: 10.1111/jtsa.12725