HAR Inference for Quantile Regression in Time Series

Jungbin Hwang
Department of Economics,
University of Connecticut

Gonzalo Valdés
Departamento de Ingeniería Industrial y de Sistemas,
Universidad de Tarapacá *†

March, 2025

Abstract

This paper develops robust inference for conditional quantile regression (QR) under unknown forms of weak dependence in time series data. We first establish fixed-smoothing asymptotic theory for QR by showing that the long-run variance (LRV) estimator for the non-smooth QR score process weakly converges to a random matrix scaled by the true LRV. Additionally, QR-Wald statistics based on the kernel LRV estimator converge to non-standard limits, while using orthonormal series LRV estimators yields standard F and t limits. For the practical implementation of our new asymptotic theory for Wald and t inference in QR, we extend heteroskedasticity and autocorrelation robust (HAR) inference for conditional mean regression to QR and apply the optimal smoothing parameter selection rule based on the Neyman-Pearson principle. Monte Carlo simulation results show that our QR-HAR procedure reduces size distortions of the HAR inference based on the conditional mean regression and the QR-HAC inference— particularly in scenarios with moderate sample sizes, strong temporal dependence, and multiple parameters in the joint null hypothesis.

JEL Classification: C12, C19, C22, C32;

Keywords: Quantile regression, heteroskedasticity and autocorrelation robust, long-run variance, alternative asymptotics, testing-optimal smoothing parameter choice

*Email: jungbin.hwang@uconn.edu; gvaldes@uta.cl

†We would like to thank Yoosoon Chang, Joon Y. Park, Ke-Li Xu, and the seminar participants at Indiana University for their helpful comments. We are also grateful to Yixiao Sun for his helpful feedback and suggestions on this paper. Jungbin Hwang also benefited from helpful conversations with Antonio Galvao and Silvia Gonçalves in the early stages of the project. Gonzalo Valdés acknowledges financial support from ANID Chile Fondecyt Regular, 1250918.

1 Introduction

Quantile regression (QR) extends traditional conditional mean regression by estimating relationships between variables at different quantiles of the dependent variable's conditional distribution (Koenker, 2005). By employing the asymmetric loss function proposed by Koenker and Basset (1978), QR offers greater flexibility than conditional mean regression and is more robust to outliers when modeling varying effects across the distribution of the dependent variable. Its application has expanded in the analysis of serially correlated outcome variables in time series economic data, as illustrated by examples in Xiao (2012) and Galvao and Yoon (2024). The goal of this paper is to provide robust inference for the conditional QR, addressing unknown forms of weak dependence in time series data.

When regression errors are serially correlated and heteroskedastic in time series data, the conventional robust standard error formulas for independent data become invalid. This issue is well recognized in the conditional mean regression models and was addressed by the pioneering work of White and Domowitz (1984), followed by Newey and West (1987) and Andrews (1991). Key innovations of these initial work include positive-definite, non-parametric estimation of the long-run variance (LRV) matrix of the regression score vector and the establishment of its heteroskedasticity and autocorrelation consistency (HAC) property. Technically, the HAC property relies on the assumption that the amount of smoothing in the nonparametric LRV estimate increases in the limiting experiment of the asymptotic approximation. The literature also offers guidance on selecting the optimal smoothing parameter for computing robust standard errors, as discussed in Andrews (1991) and Newey and West (1994).

More recently, there has been a breakthrough in the time series literature, which introduced an alternative asymptotic approximation for time series robust standard errors (Lazarus et al, 2018). This new type of asymptotic theory and the corresponding inferential methods, known as HAR inference, are based on fixed-smoothing asymptotics, which were initially developed by Kiefer and Vogelsang (2002, 2005), Müller (2007), and Sun et al. (2008) in time series econometrics. In contrast to the conventional increasing-smoothing asymptotics in the HAC approach, the alternative asymptotic theory assumes that the amount of smoothing in the non-parametric LRV estimator is fixed as the sample size grows. As a result, the LRV estimate weakly converges to a random matrix scaled by the true LRV, which contrasts with the HAC approach. The corresponding HAR-based inference show numerical improvements over HAC-type inference methods because the fixed-smoothing asymptotics can automatically provide second-order corrected inference under conventional increasing-smoothing asymptotics, see Sun et al. (2008), Sun (2014a), and Lazarus et al. (2021), for more details. The HAR approach is also directly connected to the low-frequency econometrics proposed by Müller and Watson (2020), which captures the long-run variability of economic data by utilizing low-frequency trigonometric weighted averages with frequency cutoffs corresponding to business cycles.

In this paper, we develop QR-HAR inference using more accurate fixed-smoothing asymptotics. Our alternative asymptotic theory applies to a broad class of HAR-LRV estimators, including existing kernel LRV estimators (Galvao and Yoon, 2023), exponentiated kernel estimators (Phillips et al., 2006), and orthonormal series (OS) LRV estimators (Phillips, 2005; Sun, 2013). The main theoretical challenge lies in establishing fixed-smoothing asymptotic theory for the LRV estimator of the non-differentiable and dependent QR score process, assuming that the amount of smoothing is fixed as the sample size grows to infinity. Andrews (1991) and Galvao and Yoon (2024) find that estimation errors in kernel LRV estimates for non-differentiable processes can be ignored in the contexts of the conditional mean regression and QR, respectively. However, their results crucially depend on increasing the amount of smoothing as the number of time periods grows, at a slower rate than the sample size. This assumption does not apply under the fixed-smoothing asymptotic framework considered in this paper.

To address these technical challenges arising from the non-smooth QR model, we formulate the HAR LRV estimator using the demeaned QR process, which effectively recenters the weight function in the LRV estimation. The recentering scheme allows for a spectral representation of the LRV estimator, enabling us to transform the sum of autocovariances in the LRV estimator into a scaled sum of weighted empirical processes, with weights determined by the zero-mean eigenfunctions. Additionally, we extend the standard empirical process theory for i.i.d. data, e.g., Van der Vaart and Wellner (1996), to the time series case involving non-i.i.d. data with unknown forms of dependence. With these results, our fixed-smoothing asymptotic theory accounts for the estimation uncertainty inherent in the non-differentiable QR score process. We also prove that the LRV estimate in QR weakly converges to a random matrix scaled by the true LRV, which aligns with existing results in the HAR literature for the conditional mean regression setting.

Building on the alternative asymptotic results for the HAR LRV estimator in QR, we show that the corresponding QR-Wald statistics weakly converge to non-standard limits. These non-standard limits are free of nuisance parameters, and their asymptotic critical values can be conveniently simulated by generating functions of standard Gaussian random vectors that depend on the level of smoothing. We further show that with the orthonormal series LRV (OS-LRV), the Wald and t-statistics admit standard F and t critical values, eliminating the need to simulate non-standard critical values in practice. Thus, our fixed-smoothing asymptotic results are comparable to the HAR inference framework used in conditional mean regression models, such as Sun (2014a, 2014b) and Lazarus et al. (2021). Given its convenience and improved finite-sample performance, we recommend using OS-LRV in HAR inference for the QR setting. We also show that our HAR QR inference can be extended to the non-smooth generalized method of moments (GMM) framework.

For the practical implementation of QR-HAR inference, selecting the smoothing parameter is required. In the case of the OS-LRV estimator, the smoothing parameter corresponds to the number of basis functions. Existing approach, such as Phillips (2005), recommends selecting the smoothing parameter to minimize the estimator's asymptotic mean squared error (AMSE). While this method is optimal for the point estimation of the LRV, it is not directly suited for hypothesis testing, which are the primary goals in time series robust QR inference. Our QR-HAR approach addresses this issue by developing an optimal smoothing selection rule that prioritizes hypothesis testing based on the classical Neyman-Pearson principle. Using second order approximations of Type I and Type II errors, we derive a closed-form formula for the testing-optimal smoothing parameter. This notion of the testing-optimal smoothing parameter in HAR inference was first introduced by Sun et al. (2008) and Sun (2011, 2013) in the conditional mean regression. See also Lazarus et al. (2018) and Lazarus et al. (2021) for other versions of the testing-optimal criterion within a non-NP framework.

Our Monte Carlo simulation results indicate that the QR-Wald inference approach in the median regression significantly reduces the empirical size distortions found in HAR inference based on the conditional mean regression. Its finite-sample performance remains robust to the time series persistence of the regression error, even when the error distribution is asymmetric or exhibits heavier tails than the Gaussian distribution. This finding confirms the advantage of time series robust QR inference, as pointed out in Xu (2021), which addresses both serial correlation and the effects of heavy-tailed errors with unbounded second moment that negatively impact the performance of the HAR conditional mean regression. Furthermore, our numerical results indicate that the QR-HAR approach reduces the size distortions of existing QR-HAC approach across various data-generating processes. The finite-sample improvements of our QR-HAR approach arise from the use of more accurate asymptotic F and t critical values, driven by alternative fixed-smoothing asymptotics and a testing-oriented smoothing parameter. We also find that the advantages of the QR-HAR approach over QR-HAC are particularly pronounced when the sample size is moderate, temporal dependence increases, and multiple parameters are included in the null hypothesis.

This paper contributes to the literature on robust inference in time series by addressing the challenges posed by serially correlated and heteroskedastic regression errors. Earlier works include Newey-West (1987), Andrews (1991), Kiefer and Vogelsang (2005), Sun et al. (2008), Sun (2014 a&b), and Lazarus et al. (2021). These studies predominantly focus on the conditional mean regressions or differentiable time series processes involving unknown parameters, with relatively little attention given to addressing serial correlation and heteroskedasticity in the non-smooth QR model. Chen et al. (2014) develop a time series robust sieve inference method for semi-nonparametric time series models. A recent work by Galvao and Yoon (2024) extends Andrews' (1991) HAC-approach to QR models in time series data. The key contribution of their theory is the formal development of increasing-smoothing asymptotic theory for long-run variance estimates of the QR score function. Specifically, they prove the consistency of their HAC estimator in QR and provide a rule for optimal smoothing parameter (bandwidth) selection, extending Andrews (1991)'s conditional mean regression framework. While their HAC standard errors can significantly reduce the size distortion problem associated with heteroskedasticity-consistent (HC) procedures in QR, the QR-HAC approach fails to deliver accurate empirical sizes, particularly when the degree of temporal dependence increases.

Our work fills a gap in the QR time series literature by developing a more accurate alternative asymptotic theory that fully integrates testing-oriented HAR inference in a non-differentiable QR setting. The HAR inference using the fixed-smoothing asymptotics is closely related to the self-normalized (SN) approach for weakly dependent data, as studied in Shao (2010). Zhou and Shao (2013) and Hoga and Schultz (2025) establish asymptotic theory for the SN inference for QR. The SN approach also leverages the random asymptotic limit of the denominators in self-normalized statistics. However, its practical implementation involves recursive estimations of QR coefficients using only subsamples, which can be computationally demanding and requires selecting a tuning parameter that determines the range of the initial subsample periods. In contrast, our HAR inference in QR requires only the computation of the time series robust Wald and t statistics and can utilize standard F and t critical values with a testing-oriented smoothing parameter selection. The literature also provides time series robust bootstrapping approaches to enhance the performance of QR, including the moving block bootstrap (MBB) method by Fitzenberger (1998) and the smoothed and tapered MBB method by Gregory et al. (2018). These methods commonly require additional tuning parameters and the simulation of bootstrap critical values, which can be also computationally costly. Our Monte Carlo experiments show that the finite-sample performance of our QR-HAR inference is comparable to the bootstrap approach in Gregory et al. (2018) when serial correlation is moderate and can outperform the bootstrap approach when serial correlation is strong.

The remainder of the paper is organized as follows: Section 2 introduces the issue of time series dependence in QR and proposes the HAR LRV matrix estimator for the QR model. Section 3 establishes the fixed-smoothing asymptotic theory for the QR-HAR LRV estimator. Building on this result, Section 4 develops nuisance-parameter-free asymptotic inferences for QR. This section also discusses a possible extension of our HAR QR inference to non-smooth GMM for quantile IV regression models and establishes a connection between our fixed-smoothing asymptotic inference and the self-normalization approach for QR. Section 5 provides practical recommendations for QR-HAR inferences, focusing on the optimal choice of the smoothing parameter. Section 6 presents the results of Monte Carlo simulations, and Section 7 concludes the paper. Tables and proofs are included in the Appendix of Section 8.

2 Quantile Regression for Weakly Dependent Data

2.1 The issue of time series dependence in QR

We consider the linear τ-th conditional QR equation:

$$y_t = X_t'\beta_0(\tau) \ + e_t(\tau) \text{ for } t \in \{1, \dots, T\},$$

and $\tau \in (0,1)$, where the population QR coefficient $\beta_0(\tau) \in \mathbb{R}^d$ is the unique minimizer of $\mathbb{E}[\rho_\tau(y_t - X_t'b)]$ with $\rho_\tau(u) = u \cdot (\tau - 1(u \leq 0)$. The QR estimator $\hat{\beta}(\tau)$ solves the following convex objective function (Koenker, 2005):

$$\hat{\beta}(\tau) = \arg\min_b \sum_{t=1}^{T} \rho_\tau(y_t - X_t'b),$$

and $\hat{\beta}(\tau)$ satisfies the following approximate first-order condition:

$$\frac{1}{T}\sum_{t=1}^{T} Z_t(\hat{\beta}(\tau)) = \frac{1}{T}\sum_{t=1}^{T} X_t(\tau - 1(y_t \leq X_t'\hat{\beta}(\tau)) = o_p\left(\frac{1}{\sqrt{T}}\right). \tag{1}$$

We assume that the QR score function $Z_t(b) := X_t(\tau - 1(y_t \leq X_t'b))$ identifies the population parameter $\beta_0(\tau)$ by having zero unconditional expected value, i.e.,

$$\mathbb{E}[Z_t(\beta_0(\tau))] = \mathbb{E}[X_t(\tau - 1(y_t \leq X_t'\beta_0(\tau))] = 0.$$

Let $q_y(\tau|X_t)$ represent the τ-quantile of the outcome variable y_t conditional on the regressor $X_t \in \mathbb{R}^d$, which satisfies

$$P(y_t \leq q_y(\tau|X_t)|X_t) - \tau = \mathbb{E}\left[\{1\left(y_t \leq q_y(\tau|X_t)\right) - \tau\}\,|X_t\right] = 0 \text{ a.s.} \tag{2}$$

When the τ-conditional quantile function is correctly specified in the QR equation, the conditional moment restriction above holds by construction, with $q_y(\tau|X_t) = X_t'\beta_0(\tau)$ almost surely. In this case, the QR coefficient $\hat{\beta}(\tau)$ estimates the effect or prediction of X_t on the true τ-conditional quantile of the outcome variable y_t. When $q_y(\tau|X_t)$ is misspecified, however, the above conditional moment restriction does not hold. Nevertheless, $X_t'\hat{\beta}(\tau)$ can still be interpreted as estimation of the best approximation of the true τ-conditional quantile of y_t, minimizing a weighted mean-squared error loss function for misspecification error (Angrist et al., 2006).

The goal of this paper is to provide robust inference for QR coefficient $\beta_0(\tau)$, addressing unknown forms of dependence in time series data $\{(Y_t, X_t')\}_{t=1}^{T}$. Throughout the paper, we assume $\sqrt{T}$-consistency of $\hat{\beta}(\tau)$

and that the following Bahadur representation holds:

$$\sqrt{T}(\hat{\beta}(\tau) - \beta_0(\tau)) = D(\tau)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t(\tau - 1(y_t \leq X_t'\beta_0(\tau)) + o_p(1) \tag{3}$$

$$\xrightarrow{d} N(0, \Sigma(\tau)) \text{ with } \Sigma(\tau) = D(\tau)^{-1}\Omega(\tau)D(\tau)^{-1}, \tag{4}$$

where $D(\tau) = \mathbb{E}[f(0|X_t)X_tX_t']$ and $f(0|X_t)$ is the conditional density of $e_t(\tau)$ evaluated at 0. Our asymptotic theory in this paper assumes a fixed value of $\tau \in (0,1)$, but it can be generalized to hold uniformly over $\tau \in [\epsilon, 1-\epsilon]$ by imposing certain regularity conditions on $q_y(\tau|X_t)$, such as Lipschitz continuity. For example, Galvao and Yoon (2024) extend the pointwise asymptotic theory for time series QR under the increasing smoothing asymptotics.

In the presence of an unknown form of weak dependency in the time series data $\{(Y_t, X_t')\}_{t=1}^{T}$, the $d \times d$ matrix $\Omega(\tau)$ is a long-run variance (LRV) matrix of the quantile score process

$$\Omega(\tau) = \sum_{j=-\infty}^{\infty} \Gamma_j(\tau) \text{ with } \Gamma_j(\tau) = Cov(Z_t, Z_{t-j}),$$

where $Z_t := Z_t(\beta_0(\tau)) = X_t(\tau - 1(y_t \leq X_t'\beta_0(\tau))$.

Let $\hat{\Sigma}(\tau) := \hat{D}(\tau)^{-1}\hat{\Omega}(\tau)\hat{D}(\tau)^{-1}$ denote an estimate for $\Sigma(\tau){:=}\ D(\tau)^{-1}\Omega(\tau)D(\tau)^{-1}$, which is a crucial component for robust Wald and t statistics in QR. When formulating $\hat{\Sigma}(\tau)$, non-parametric estimations for $D(\tau)$ and $\Omega(\tau)$ are required. The term $D(\tau)$ captures the impact of heteroskedasticity and has been extensively studied in the QR literature. Specifically, Kato (2012) proves that Powell's sandwich estimator, $\hat{D}(\tau) = (Tl_T)^{-1}\sum_{t=1}^{T} K((y_t - X_t'\hat{\beta}(\tau))/l_T)(X_tX_t')$, where $K(u) := 2^{-1}1(|u| \leq 1)$ denotes the uniform kernel function, is consistent and asymptotically normal under a broad range of data-generating processes, including both i.i.d. and time series settings. It is noteworthy that the formulation of $\hat{D}(\tau)$ remains unchanged for dependent time series data. A more detailed implementation of $\hat{D}(\tau)$, including the bandwidth parameter, l_T is provided in subsection 8.1 of Appendix.

In contrast to $\hat{D}(\tau)$, the formulation of $\hat{\Omega}(\tau)$ involves a non-trivial difference from the i.i.d. case. This is because the presence of an unknown form of time series dependence introduces infinite summations of autocovariance for the non-differentiable QR score process $\{Z_t\} = \{m_t(\tau)X_t\}$, where $m_t(\tau) := \tau - 1(e_t(\tau) \leq 0)$, in $\Omega(\tau)$. This dependence not only arises from the time series regressor X_t but can also be induced by the autocorrelation in the QR error $e_t(\tau)$ via its quantile score function $m_t(\tau)$. The serial correlation in $m_t(\tau)$ captures the temporal dependence of tail events at the τ-quantile. It can arise even if the true QR is correctly specified, as in (2). For example, suppose the true data-generating process (DGP) of y_t follows a simple AR(1) process:

$$y_t = \mu_y + \rho y_{t-1} + \epsilon_t, \tag{5}$$

where $|\rho| < 1$, ϵ_t is independent of y_{t-1}. In this case, the true conditional quantile function is correctly specified with $q_y(\tau|X_t) = X_t'\beta_0(\tau)$, where $X_t = (1, y_{t-1})$, $\beta_0(\tau) = (F_\epsilon^{-1}(\tau) + \mu_y, \rho)'$, and $F_\epsilon^{-1}(\cdot)$ denotes the inverse of the cumulative distribution function (CDF) of ϵ_t. Also, it is not difficult to check that the τ-quantile error $e_t(\tau) = \epsilon_t - F_\epsilon^{-1}(\tau)$ satisfies

$$\mathbb{E}[m_t(\tau)|X_t] = \mathbb{E}[\tau - 1(e_t(\tau) \leq 0)|X_t] = 0 \text{ a.s.} \tag{6}$$

The process $m_t(\tau)$ can exhibit serial correlation if the error process ϵ_t is serially correlated. For instance, Galvao and Yoon (2024) numerically illustrate that the quantile autocorrelation of $m_t(\tau)$ strictly increases with respect to ρ in the Gaussian process case of $\{\epsilon_t\}$ (or $\{e_t(\tau)\}$). The serial correlations embodied in $m_t(\tau)$ and X_t imply that

$$\Gamma_j(\tau) = Cov(Z_t, Z_{t-j}) = \mathbb{E}[X_t X_{t-j}]\mathbb{E}[m_t(\tau)m_{t-j}(\tau)] \neq 0 \text{ for } j \neq 0.$$

Consequently, conventional heteroskedasticity-robust standard error estimators in QR, which assume independent error terms—such as those proposed by Angrist et al. (2006, p. 551)—do not apply to time series data, as they rely on the assumption $\Omega(\tau) = \Gamma_0(\tau)$, which does not hold in the presence of serial correlation. Beyond the simple AR(1) example, the issues of the serial correlation in $m_t(\tau)$ can be arisen in many other empirical settings, including h-period-ahead quantile prediction, as in Adrian et al. (2019). In this case, $y_t = Y_{t+h}$ for $h > 0$, which naturally induces serial correlation in the QR regression error $e_{t+h}(\tau)$.

We also point out that even if the error process in the conditional mean regression is serially uncorrelated, it does not necessarily follow that $\Gamma_j(\tau) = 0$ for $j \neq 0$. For example, consider the following non-AR(1) DGP for y_t :

$$y_t = \beta_{0,0} + \beta_{0,1}x_t + \epsilon_t,$$

where $\{x_t\}$ is a strictly stationary and serially correlated time series that is independent of $\{\epsilon_t\}$. Let $\mathcal{F}_{t-1}$ denote the σ-algebra generated by the process $\{x_{t-1}, \epsilon_{t-1}, x_{t-2}, \epsilon_{t-2}, \ldots\}$, and suppose $\epsilon_t = \sigma_t u_t$, where $\{u_t\}$ is an i.i.d innovation series and σ_t is $\mathcal{F}_{t-1}$-measurable positive random variable following an ARCH or GARCH-type process. In this case, ϵ_t satisfies the martingale difference sequence (m.d.s.) condition, i.e., $\mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = 0$ a.s., and thus $E[\epsilon_t\epsilon_{t-j}] = 0$. Moreover, the true conditional quantile function can be correctly specified as $q_y(\tau|X_t) = X_t'\beta_0(\tau)$, where $X_t = (1, x_t)$, $\beta_0(\tau) = (F_\epsilon^{-1}(\tau) + \beta_{0,0}, \beta_{0,1})'$. However, $m_t(\tau) = \tau - 1(\sigma_t u_t - F_\epsilon^{-1}(\tau) \leq 0)$, which is the nonlinear transformation of $\epsilon_t = \sigma_t u_t$, can exhibit serial correlation for $\tau \in (0, 1)$ such that $F_\epsilon^{-1}(\tau) \neq 0$, due to the serial dependence in the conditional heteroskedasticity process σ_t. To ensure that $\Gamma_j(\tau) = 0$ for all $j \neq 0$, we shall impose the m.d.s. assumption of $\mathbb{E}[m_t(\tau)|\mathcal{F}_{t-1}] = 0$ a.s. This condition, however, is more restrictive than assuming a correctly specified conditional quantile function as in (6) and essentially rules out the presence of serially correlated conditional heteroskedasticity in ϵ_t.

2.2 HAR LRV estimation in QR

For simplicity of the notation, we drop the dependence on the given quantile level τ in $e_t(\tau)$ and use e_t hereafter and denote $e_t(b) := y_t - X_t'b$ and $\hat{Z}_t = X_t(\tau - 1(\hat{e}_t \leq 0))$ with $\hat{e}_t := e_t(\hat{\beta}(\tau))$. The time series corrected standard errors and robust inference begin by formulating heteroskedasticity and autocorrelation robust (HAR) LRV matrix estimator $\hat{\Omega}(\tau)$. In this paper, we study the following general class of the quadratic HAR variance estimators. The LRV estimator takes the form:

$$\hat{\Omega}_h(\tau) := \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h\left(\frac{t}{T}, \frac{s}{T}\right)\hat{Z}_t^c \hat{Z}_s^{c\prime}, \tag{7}$$

where $\hat{Z}_t^c = \hat{Z}_t - T^{-1}\sum_{s=1}^{T}\hat{Z}_s$, and $Q_h(\cdot,\cdot)$ is a symmetric weighting function which uses a smoothing parameter h. By construction, our formulation of $\hat{\Omega}_h(\tau)$ depends on h which is the level of smoothing for various types of LRV estimators. Examples of $\hat{\Omega}_h(\tau)$ include a popular class of conventional kernel LRV estimators with $Q_h(r,s) = k((r-s)/b)$ and $h = 1/b$, as studied in Galvao and Yoon (2024). The kernel LRV estimators include the popular Newey-West (1987) HAC estimator with the Bartlett kernel $k(x) = 1(|x| \leq 1)\cdot(1-|x|)$, as well as those utilize second-order kernels such as Parzen and Quadratic Spectral functions, e.g., Andrews (1991). When $Q_h(r,s) = k^{\rho}(r-s)$ with $h = \sqrt{\rho}$, $\hat{\Omega}_h(\tau)$ belongs to a class of exponentiated kernel LRV estimators in Phillips et al. (2006).

Another important class is a class of orthonormal series LRV (OS-LRV) estimators, which is proposed by Phillips (2005), Sun (2013), and Müller and Watson (2014). The OS-LRV estimator takes $h = K$ and sets the weight function equal to

$$Q_h(r,s) = \frac{1}{K}\sum_{k=1}^{K}\Phi_k(r)\Phi_k(s), \tag{8}$$

where $\{\Phi_k(r)\}_{k=1}^{K}$ is a set of orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1 \Phi_k(r)\,dr = 0$. In this paper, we employ the following set of finite basis functions:

$$\{\Phi_k(r)\}_{k=1}^{K} = \{\Phi_{2k-1}(r) = \sqrt{2}\sin(2\pi kr), \Phi_{2k}(r) = \sqrt{2}\cos(2\pi kr), k \in \{1,2,\ldots,K/2\}\}, \tag{9}$$

where K is an even number. From this construction, the OS-LRV is an equal-weighted periodogram (EWP) estimator that utilizes the first $K/2$ periodograms, e.g., Sun (2013) and Lazarus et al. (2019), which corresponds to an estimator of the scaled spectral density at zero.

Let $\hat{\Omega}_h^u(\tau)$ be an uncentered version of (7) that replaces $\hat{Z}_t^c$ with $\hat{Z}_t$. Considering the kernel weighting function $Q_h(r,s) = k((r-s)/b)$, with $S_T = Tb$ and $h = 1/b$, Galvao and Yoon (2023, Theorem 4.2-(a)) show that

$$\hat{\Omega}_h^u(\tau) = \tilde{\Omega}_h^u(\tau) + O_p\left(\frac{S_T}{\sqrt{T}}\right) \tag{10}$$

holds, as $S_T, T \to \infty$, where $\tilde{\Omega}_h^u(\tau)$ denotes the infeasible version of $\hat{\Omega}_h^u(\tau)$ that replaces $\hat{Z}_t$ with Z_t. Based on this result, Galvao and Yoon (2024) further establish the consistency of $\hat{\Omega}_h^u(\tau)$, extending Andrews (1991) in the conditional mean regression setting, i.e., $\hat{\Omega}_h^u(\tau) \xrightarrow{p} \Omega(\tau)$. Their key rate condition is $S_T/\sqrt{T} = o(1)$, which can be equivalently expressed as $b = o(T^{-1/2})$ and thus falls within the conventional increasing-smoothing (or small-b) asymptotics. The conventional asymptotics, elegant and convenient though it may be, it completely ignore the estimation uncertainty in $\hat{\Omega}_h^u(\tau)$ which is the key component of the time series robust inference for QR. This issue has been well-recognized in the literature on HAR inference for conditional mean regression, e.g., Kiefer and Vogelsang (2002, 2005), Müller (2007), and Sun et al. (2008, 2014a), and Hwang an Sun (2018), which naturally motivates a more accurate fixed-smoothing asymptotic theory in QR, developed in this paper.

In (10), the term $O_p(S_T/\sqrt{T})$ captures the estimation uncertainty of $Z_t(\hat{\beta}(\tau))$ in $\hat{\Omega}_h^u(\tau)$, with a stochastic order of $O_p(S_T/\sqrt{T}) = O_p(\sqrt{T}b)$. This term converges to zero in probability under the increasing-smoothing (or small-b) asymptotics with rate $b = o(T^{-1/2})$. However, under our fixed-smoothing (fixed-b) asymptotic framework, the $O_p(\sqrt{T}b)$ term does not degenerate to zero in probability since $b = O(1)$, rendering the result in (10) inapplicable to our setting.

To address this challenge, our asymptotic analysis focuses on the demeaned $\hat{\Omega}_h(\tau)$ instead of $\hat{\Omega}_h^u(\tau)$. In the mean linear regression model, the estimated score process is $\hat{Z}_t = X_t\hat{e}_t$, whose sample mean $T^{-1}\sum_{s=1}^T \hat{Z}_s$ is always equal to zero. Thus, the use of the original $\hat{Z}_t$ in (7) does not change $\hat{\Omega}_h(\tau)$ from $\hat{\Omega}_h^u(\tau)$. In contrast, the non-linear and non-smooth QR score process in our setting yields a non-zero small-order term $T^{-1}\sum_{s=1}^T \hat{Z}_s$, whose stochastic order of magnitude is $o_p(T^{-1/2})$. As a result, the use of demeaned $\hat{Z}_t^c$ in $\hat{\Omega}_h(\tau)$ can create a difference from $\hat{\Omega}_h^u(\tau)$ in finite samples. We note that the issue of a nonzero term $T^{-1}\sum_{s=1}^T \hat{Z}_s$ becomes more important if the non-smooth moment condition is over-identified. See, for example, Hong and Li (2023), in the context of misspecified non-smooth moment conditions. In our exactly identified QR setting, we employ the demeaned term $\hat{Z}_t^c$ in LRV estimates to prove that the estimation uncertainty inherent in $\hat{Z}_t$ (via $\hat{\beta}(\tau)$) is properly controlled under the fixed-smoothing asymptotic framework. This process is discussed in more detail in the next section, which presents the main theoretical contribution of this paper.

3 Fixed-smoothing Asymptotic Theory for Quantile Regression

To develop fixed-smoothing asymptotic theory for $\hat{\Omega}_h(\tau)$, we start by reexpressing $\hat{\Omega}_h(\tau)$ as

$$\hat{\Omega}_h(\tau) = \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T Q_{T,h}^*\left(\frac{t}{T}, \frac{s}{T}\right)\hat{Z}_t\hat{Z}_s', \tag{11}$$

where $Q^*_{T,h}(r,s) = Q_h(r,s) - T^{-1}\sum_{\tilde{r}=1}^{T} Q_h(\tilde{r}/T, s) - T^{-1}\sum_{\tilde{s}=1}^{T} Q_h(r, \tilde{s}/T) + T^{-1}\sum_{\tilde{r}=1}^{T}\sum_{\tilde{s}=1}^{T} Q_h(\tilde{r}/T, \tilde{s}/T)$. Let $\tilde{\Omega}_h(\tau)$ be an infeasible version of $\hat{\Omega}_h(\tau)$, which replaces the plugged-in estimate for $\hat{Z}_t$ in $\hat{\Omega}(\tau)$ with Z_t, i.e.,

$$\tilde{\Omega}_h(\tau) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q^*_{T,h}\left(\frac{t}{T}, \frac{s}{T}\right) Z_t Z_s'. \tag{12}$$

The key step in our asymptotic approximation for $\hat{\Omega}_h(\tau)$ is to show that

$$\hat{\Omega}_h(\tau) = \tilde{\Omega}_h(\tau) + o_p(1) \tag{13}$$

holds, as $T \to \infty$ such that the smoothing parameter h is fixed. To achieve this goal, we introduce an asymptotically equivalent version of (11):

$$\hat{\Omega}^*_h(\tau) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q^*_h\left(\frac{t}{T}, \frac{s}{T}\right) \hat{Z}_t \hat{Z}_s',$$

where $Q^*_h(r,s) := \lim_{T\to\infty} Q^*_{T,h}(r,s)$ is the limit of the centered weight function, given by:

$$Q^*_h(r,s) = Q_h(r,s) - \int_0^1 Q_h(\tilde{r}, s) d\tilde{r} - \int_0^1 Q_h(r, \tilde{s}) d\tilde{s} + \int_0^1\int_0^1 Q_h(\tilde{r}, \tilde{s}) d\tilde{r} d\tilde{s}. \tag{14}$$

In the proof of Theorem 1 below, we show that $\hat{\Omega}^*_h(\tau) - \hat{\Omega}_h(\tau) = o_p(1)$ holds for any fixed h. This allows us to focus on $\hat{\Omega}^*_h(\tau)$ with the following representation of the centered weighting function $Q^*_h(r,s)$:

$$Q^*_h(r,s) = \sum_{k=1}^{\infty} \lambda_k \Phi_k(r) \Phi_k(s), \tag{15}$$

where $\{\Phi_k(\cdot)\}_{k=1}^{\infty}$ is a sequence of continuously differentiable orthonormal basis functions over $[0,1]$. For notational simplicity, we suppress the dependencies of h on $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\Phi_k(\cdot)\}_{k=1}^{\infty}$. The boundedness of $Q^*_h(r,s)$ implies that $\Phi_k(\cdot)$ are uniformly bounded over $[0,1]$ for all $k \in \mathbb{N}$. Additionally, the orthonormal property, $\Phi_k(\cdot)$ ensures that

$$\int_0^1 \Phi_k(r) dr = 0 \text{ and } \int_0^1\int_0^1 \Phi_j(s)\Phi_k(r) ds dr = 1(j = k)$$

for all $j, k \in \mathbb{N}$. As a result, $\sum_{k=1}^{\infty} \lambda_k = \int_0^1 Q^*_h(r,r)\,dr$ is finite and is uniformly bounded over h. Also, the series on the right-hand side of (15) converges to $Q^*_h(r,s)$ absolutely and uniformly over $(r,s) \in [0,1]\times[0,1]$, which allows us to re-express $\hat{\Omega}^*_h(\tau)$ in the following form:

$$\hat{\Omega}^*_h(\tau) = \sum_{k=1}^{\infty} \lambda_k \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right)\hat{Z}_t\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T} \Phi_k\left(\frac{s}{T}\right)\hat{Z}_s\right)'. \tag{16}$$

Since $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\Phi_k(\cdot)\}_{k=1}^{\infty}$ are the eigenvalues and eigenfunctions of the bivariate positive definite function $Q^*_h(r,s)$, this representation constitutes a spectral decomposition of the compact Fredholm operator with kernel $Q^*_h(r,s)$, see Knessl and Keller (1991) and Sun (2014b) for more details. We note that the eigen decomposition in (15) is used solely for theoretical proof to establish the fixed-smoothing limit of $\hat{\Omega}_h(\tau)$ in (7). As shown in the next section, computing $\hat{\Omega}_h^*(\tau)$ with (16) is not necessary in practice to apply our fixed-smoothing asymptotic theory to HAR inference.

Taking advantage of the expression in (16), we can address the estimation uncertainties embodied in $\hat{\Omega}_h^*(\tau)$. Consider the following component using the k-th order basis function $\Phi_k(\cdot)$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\hat{Z}_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)X_t(\tau - 1(\hat{e}_t \leq 0)) \tag{17}$$

which is the weighted average of the estimated QR score process $\hat{Z}_t$. Due to the non-differentiability of $Z_t(b) = X_t(\tau - 1(y_t \leq X_t'b))$ with respect to b, the standard Taylor expansion for the right-hand side of (17) is not possible. To deal with this challenge under fixed-smoothing asymptotics, we impose the following assumptions.

Assumption 1. *(i) For kernel LRV estimators, the positive definite kernel function $k(\cdot) \in [-1,1]$ satisfies the following conditions: For any $b \in (0,1]$ and $\rho \in [1,\infty)$, $k_b(x) = k(x/b)$ and $k^\rho(x) = k^\rho(x)$ are symmetric, continuous, piecewise monotonic, and piecewise continuously differentiable functions on $[-1,1]$ such that $\int_{-\infty}^{\infty} k^2(x) < \infty$. (ii) For the OS-LRV variance estimator, the basis functions $\Phi_k(x)$ are piecewise monotonic, continuously differentiable, and orthonormal in $L^2[0,1]$, and $\int_0^1 \Phi_k(x)\,dx = 0$.*

Assumption 2. *The vector process $\{(Y_t, X_t')'\}_{t=1}^T$ is strictly stationary and α-mixing, where the mixing coefficient $\alpha[k]$ satisfies $\alpha[k] \leq \exp(-a_0 k)$ for some $a_0 > 0$.*

Assumption 3. *Let $f(e|x)$ be the conditional density of e_t given $X_t = x$. i) For each x, $f(e|x)$ is twice continuously differentiable with respect to e and satisfies the following conditions: $f(0|x) > 0$ and $|f(e|x)|, |f'(e|x)|, |f''(e|x)| \leq C$ for some positive constant C which does not depend on x and e. ii) The first order condition in (1) and the Bahadur representation in (3) hold. iii) The matrix $\Omega(\tau)$ is non-singular.*

Assumption 4. *i) The zero mean process $\{Z_t\}_{t=1}^T$ with $Z_t = X_t(\tau - 1(e_t \leq 0))$ satisfies that $\sum_{j=-\infty}^{\infty} ||\mathbb{E}[Z_t Z_{t-j}]|| < \infty$ and $\sum_{j=-\infty}^{\infty} |j|^{q_0}||\mathbb{E}[Z_t Z_{t-j}]|| < \infty$ for some $q_0 \geq 1$. ii) There exists a constant $\Delta > 0$ such that $||X||_\infty := \max_{1\leq t\leq T}||X_t|| \leq \Delta T^{1/5}$. iii) $\mathbb{E}[||X_t||^{4\nu}] \leq \infty$ for some $\nu > 1$, and, with $X_t = (X_{1,t}, \ldots, X_{d,t})'$, $\max_{v_1,v_2\in\{1,\ldots,4\}} \mathbb{E}[|X_{a,t}^{v_1} X_{b,t+s}^{v_2}|] < \infty$ for all $a, b \in \{1,\ldots,d\}$ and $s \in \{0, \pm 1, \pm 2, \ldots\}$.*

Assumption 5. *As $T \to \infty$, $T^{-1/2}\sum_{t=1}^T \Phi_k(t/T)Z_t$ converges weakly to a continuous distribution, jointly over $k \in \{0,1,...,J\}$ such that*

$$\begin{aligned} &P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)Z_t \leq x \text{ for } k = 0,1,...,J\right) \\ &= P\left(\Omega^{1/2}(\tau)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)u_t \leq x \text{ for } k = 0,1,...,J\right) + o(1), \end{aligned}$$

for every fixed $J \in \mathbb{N}$, $x \in \mathbb{R}^d$ *where* $\Phi_0(\cdot) = 1$, $u_t \overset{i.i.d.}{\sim} N(0, I_d)$, *and* $\Omega^{1/2}(\tau)$ *is a matrix square root of* $\Omega(\tau)$ *such that* $\Omega^{1/2}(\tau)\Omega^{1/2}(\tau)' = \Omega(\tau)$.

Assumption 1 is on the kernel function and basis functions that construct $Q_h(r, s)$ in $\hat{\Omega}_h(\tau)$. It consists of the same conditions as the smooth moment condition outlined in Sun (2014b), implying that all aforementioned HAR variance estimator can be applied to the QR setting. Additionally, Assumption 1 ensures the unified representation in (16) for all types of HAR variance estimators we consider.

The α-mixing and decaying rate conditions in Assumption 2 implies that $\alpha[k] \leq Ba^k$ for some $a \in (0, 1)$ and $B > 0$. As a result, we have that $\sum_{k=1}^{\infty}(\alpha[k])^{1/2} < \infty$, which is used to bound the variance of the empirical process $\mathbb{G}_{k,T}(b)$ defined below. Examples for Assumption 2 include a linear time series process for $e_t = \sum_{j=0}^{\infty} a_j v_{t-j}$, where the coefficient a_j shrink to zero exponentially fast, and $\{v_t\}$ are independent with a finite second moment. A sufficient condition for α-mixing requires v_t to be a continuous random variable with a smooth density, which excludes the case where v_i is a Bernoulli-type random variable. For more detailed discussions on mixing processes, see Andrews (1984) and Tuan and Lanh (1985).

Assumption 3 consists of mild conditions similar to Assumption 2 in Galvao and Yoon (2024). It includes the standard Bahadur representation for time series QR and the positive definite LRV matrix. Assumption 4-i) imposes on bounds on the infinite summations of the autocovaraince matrices, which is comparable to the smoothness of spectral density for the QR score process Z_i. The second part of Assumption 4 is similar to Assumption 3 in Galvao and Yoon (2024), and imposes standard boundedness conditions on the norm and maximum of the regressors.

We note that, in comparison to the increasing-smoothing asymptotic framework utilized by HAC approach, our fixed-smoothing asymptotics approach demands fewer technical conditions in Assumptions 3 and 4. Specifically, we do not impose the strict requirements on the joint conditional density of the QR error, the higher-order cumulants of the quantile scores process, or summability conditions for the derivatives of the covariance matrices, as outlined by Assumptions 3 and 5 in Galvao and Yoon (2024). This indicates the advantage of our fixed-smoothing asymptotic approach, which can be applied to a broader range of data generation processes in time series QR.

Our last condition in Assumption 5 is to impose a joint Central Limit Theorem (CLT) condition on the weighted score process $T^{-1/2}\sum_{t=1}^{T}\Phi_k(t/T)\, Z_t$ for $k \in \{1, \ldots, J\}$ and for every fixed $J \in \mathbb{N}$. It implies that $T^{-1/2}\sum_{t=1}^{T}\Phi_k(t/T)Z_t = O_p(1)$ holds uniformly over $k \in \{1, \ldots, J\}$. The assumption also guarantees the Gaussian approximation of the HAR variance estimator, which is the key result of our fixed-smoothing asymptotic theory for $\hat{\Omega}_h(\tau)$. For completeness, we provide some primitive sufficient conditions for the CLT condition in subsection 8.3 of Appendix.

Let $\mathcal{N}_{\epsilon_T}(\beta_0(\tau)) = \{b : ||b - \beta_0(\tau)|| \leq \epsilon_T\}$ be a shrinking neighborhood of the true parameter $\beta_0(\tau)$ for a positive sequence ϵ_T that converges to zero at $\sqrt{T}$-rate. We define an empirical process for $Z_t(b)$ over $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$ which is weighted by the orthonormal basis function $\Phi_k(\cdot)$:

$$\mathbb{G}_{k,T}(b) := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right)(Z_t(b) - \mathbb{E}[Z_t(b)])$$

for $k \in \{0\} \cup \mathbb{N}$. When $k = 0$, we let $\Phi_k(\cdot) = 1$ so that corresponding $\mathbb{G}_{k,T}(b)$ becomes the standard empirical process. Under Assumptions 1–4, Lemma 4 in subsection 8.4 of Appendix proves that

$$\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\mathbb{G}_{k,T}(b) - \mathbb{G}_{k,T}(\beta_0(\tau))|| = o_p(1) \tag{18}$$

holds uniformly over $k \in \{0\} \cup \mathbb{N}$. This result extends the standard empirical process theory for i.i.d. data, e.g., Van Der Varrt and Wellner (1996), to a sequence of time series empirical processes with weights given by $\Phi_k(\cdot)$. The result provides the essential technical step to formally prove the result in (13). Specifically, we control the uncertainty of $\hat{\beta}(\tau)$ by considering the following relation:

$$\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right)(Z_t(\beta_0(\tau)) - Z_t(\hat{\beta}(\tau))) \right\| \tag{19}$$

$$\leq \sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\mathbb{G}_{k,T}(b) - \mathbb{G}_{k,T}(\beta_0(\tau))|| + \left(\frac{1}{T} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right)\right) \cdot \left(\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} \sqrt{T}||\mathbb{E}[Z_t(b)]||\right). \tag{20}$$

By virtue of (18), the first term on the right-hand side of (20) shrinks to zeros in probability uniformly over k. Additionally, the standard QR assumptions imposed in Assumption 3, including the absolute continuity of $e_t = e_t(\tau)$, bounded moments of X_i guarantee the smoothness of the expected value in the second term of (20). Taking advantage of these results, we can apply the standard Taylor expansion to $\mathbb{E}[Z_t(b)]$ and obtain the stochastic boundedness, i.e., $\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau)),} \sqrt{T} \cdot ||\mathbb{E}[Z_t(b)]|| = O_p(1)$.

It is important to point out that the zero mean and uniform boundedness properties of basis functions $\{\Phi_k(\cdot)\}_{k=1}^{\infty}$ guarantee that the second term in (20) shrinks to zero, i.e., $T^{-1} \sum_{t=1}^{T} \Phi_k(t/T) = o(1)$, uniformly over $k \in \mathbb{N}$. The (asymptotic) zero mean property of $\{\Phi_k(\cdot)\}_{k=1}^{\infty}$ stems from the demeaned score process $\hat{Z}_t^c$ in our LRV estimator formula (7). This property is then carried over to the recentering of the original weight function $Q_h(\cdot, \cdot)$ to $Q_h^*(\cdot, \cdot)$ presented in (14) and (15). As a result, as $T \to \infty$, we have that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) \underbrace{Z_t(\hat{\beta}(\tau))}_{=\hat{Z}_t} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) Z_t + o_p(1),$$

where the $o_p(1)$ term does not depend on $k \in \mathbb{N}$. Together with $\sum_{k=1}^{\infty} \lambda_k = O(1)$ and the uniform boundedness of $\Phi_k(\cdot)$ over $k \in \mathbb{N}$, this result allows us to control all estimation errors captured in (16), i.e., as $T \to \infty$ such that h is fixed,

$$\hat{\Omega}_h^*(\tau) = \tilde{\Omega}_h^*(\tau) + o_p(1),$$

where $\tilde{\Omega}_h^*(\tau)$ is the infeasible version of $\hat{\Omega}_h^*(\tau)$, given by

$$\tilde{\Omega}_h^*(\tau) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h^*\left(\frac{t}{T}, \frac{s}{T}\right) Z_t Z_s'. \tag{21}$$

Building on these findings, Theorem 1 below formally proves the asymptotic equivalence in (13). We also show that the LRV estimate weakly converges to a random limit, which extends the fixed-smoothing asymptotic theory in the conditional mean regression, such as Kiefer and Vogelsang (2005), Sun (2014 a&b), for the QR setting.

Theorem 1. *Suppose Assumptions 1–5 hold. Then, as $T \to \infty$ such that h is fixed, the asymptotic equivalence result in (13) holds. Also, we have that*

$$\hat{\Omega}_h(\tau) \xrightarrow{d} \Omega_{h,\infty}(\tau) := \Omega^{1/2}(\tau)\mathbb{S}_{h,\infty}\Omega^{1/2}(\tau)', \tag{22}$$

where

$$\mathbb{S}_{h,\infty} = \sum_{j=1}^{\infty} \lambda_j \mathbb{Z}_j \mathbb{Z}_j' \text{ with } \mathbb{Z}_j \overset{i.i.d.}{\sim} N(0, I_d). \tag{23}$$

The proof of Theorem 1 in subsection 8.5 of Appendix. It shows that the recentered QR score process $\hat{Z}_t^c$ in $\hat{\Omega}_h(\tau)$ effectively removes the estimation uncertainty of $\hat{\beta}(\tau)$ embodied in $\hat{\Omega}_h(\tau)$. Consequently, the fixed-smoothing limit of $\hat{\Omega}_h(\tau)$ can be obtained by investigating its infeasible counterparts $\tilde{\Omega}_h(\tau)$ and $\tilde{\Omega}_h^*(\tau)$, which are defined in (12) and (21), respectively. Specifically, the weak convergence result in Lemma 1-(d) of Sun (2014b) can be applied to $\tilde{\Omega}_h^*(\tau)$, yielding

$$\tilde{\Omega}_h^*(\tau) \xrightarrow{d} \Omega^{1/2}(\tau)\left(\int_0^1\int_0^1 Q_h^*(r,s)\, dW_d(r)dW_d'(s)\right)\Omega^{1/2}(\tau)'. \tag{24}$$

The series representation in (15) then leads us to an equivalent representation of the weak convergence limit in (24):

$$\Omega^{1/2}(\tau)\sum_{j=1}^{\infty}\lambda_j\left(\int_0^1 \Phi_j(r)\, dW_d(r)\right)\left(\int_0^1 \Phi_j(s)\, dW_d(s)\right)'\Omega^{1/2}(\tau)', \tag{25}$$

where $W_d(\cdot)$ is d-dimensional standard Brownian motion. Utilizing the zero mean and orthonormal properties of $\{\Phi_k(\cdot)\}_{k=1}^{\infty}$, we have then $\int_0^1 \Phi_k(r)\, dW_d(r) \overset{d}{=} \mathbb{Z}_k$ with $\mathbb{Z}_k \overset{\text{i.i.d.}}{\sim} N(0, I_d)$ and obtain the weak convergence limit in (23).

4 HAR Inference for Quantile Regression

4.1 Fixed-smoothing asymptotic inference for QR

Building upon the fixed-smoothing asymptotic theory developed in the previous section, our next goal is to provide more accurate HAR-based asymptotic inferences for QR model. We consider the following standard Wald statistic for testing $H_0 : R\beta_0(\tau) = r \in \mathbb{R}^p$:

$$\mathbb{W}_T(\tau) := \left(R\hat{\beta}(\tau) - r\right)' \left(R\hat{\Sigma}(\tau)R'/T\right)^{-1} \left(R\hat{\beta}(\tau) - r\right), \tag{26}$$

where R is a $p \times d$ matrix with $p \leq d$ and full rank, and $\hat{\Sigma}(\tau) = \hat{D}(\tau)^{-1}\hat{\Omega}(\tau)\hat{D}(\tau)^{-1}$. For $p = 1$ and for one-sided alternatives, the t statistic can be defined as

$$\mathbb{T}_T(\tau) := \frac{R\hat{\beta}(\tau) - r}{\sqrt{R\hat{\Sigma}(\tau)R'/T}}. \tag{27}$$

The result in Theorem 1 shows that $\hat{\Omega}_h(\tau)$ in $\hat{\Sigma}(\tau)$ converges in distribution to the random matrix $\Omega_{h,\infty}(\tau)$, which is a scaled infinite mixture of quadratic Gaussian vectors. Our next asymptotic theory establishes non-standard limiting distributions for Wald and t statistics under fixed-smoothing asymptotics.

Theorem 2. *Let $\mathbb{S}_{h,\infty}^{[p]} = \sum_{j=1}^{\infty} \lambda_j \mathbb{Z}_{p,j}\mathbb{Z}_{p,j}'$, where $\mathbb{Z}_{p,j} := \int_0^1 \Phi_j(r)\, dW_p(r) \overset{i.i.d.}{\sim} N(0, I_p)$ and $\mathbb{Z}_p := W_p(1) \sim N(0, I_p)$, which is independent of $\mathbb{S}_{h,\infty}^{[p]}$. Under Assumptions 1–5, and with $\hat{D}(\tau) \xrightarrow{p} D(\tau)$, the following statements hold as $T \to \infty$ such that h fixed:*

(a) $\mathbb{W}_T(\tau) \xrightarrow{d} \mathbb{W}_\infty := \mathbb{Z}_p'[\mathbb{S}_{h,\infty}^{[p]}]^{-1}\mathbb{Z}_p$ and $\mathbb{T}_T(\tau) \xrightarrow{d} \mathbb{T}_\infty = \mathbb{Z}_p/\sqrt{\mathbb{S}_{h,\infty}^{[p]}}$ with $p = 1$.

(b) For the OS-LRV case, $\mathbb{W}_\infty$ and $\mathbb{T}_\infty$ in (a) satisfy that $\mathbb{W}_\infty \overset{d}{=} (pK)/(K-p+1)\mathcal{F}_{K-p+1}$ and $\mathbb{T}_\infty \overset{d}{=} \mathcal{T}_K$ with $p = 1$, where $\mathcal{F}_{p,K-p+1}$ is F-distribution with degrees of freedom $(p, K-p+1)$, and $\mathcal{T}_K$ is t-distribution with degree of freedom K.

Theorem 2-(a) shows that the fixed-smoothing asymptotic distributions for $\mathbb{W}_T(\tau)$ and $\mathbb{T}_T(\tau)$ do not coincide with the standard chi-square and Gaussian limits but instead yield non-standard limits. The primary source of these non-standard limits is the random denominator $\mathbb{S}_{h,\infty}^{[p]}$, which depends on the choice of the weight function $Q_h(r, s)$ and the smoothing parameter h, as reflected through $\{\lambda_j\}_{j=1}^{\infty}$. Nevertheless, the fixed-smoothing limits, $\mathbb{W}_\infty$ and $\mathbb{T}_\infty$, are free of nuisance parameters such as $D(\tau)$ and $\Omega(\tau)$. This nuisance-free property enables the simulation of their critical values in practice. Specifically, with $u_t \overset{i.i.d.}{\sim} N(0, I_p)$ and $\bar{u}_B = B^{-1}\sum_{t=1}^{B} u_t$, define

$$\begin{aligned}\mathbb{S}_{h,B}^{[p]} &= \frac{1}{B}\sum_{t=1}^{B}\sum_{s=1}^{B} Q_h\left(\frac{t}{B}, \frac{s}{B}\right)(u_t - \bar{u}_B)(u_s - \bar{u}_B)' \\ &= \frac{1}{B}\sum_{t=1}^{B}\sum_{s=1}^{B} Q_{B,h}^{*}\left(\frac{t}{B}, \frac{s}{B}\right) u_t u_s'.\end{aligned} \tag{28}$$

From the equivalent representation of the random limit

$$\mathbb{S}_{h,\infty}^{[p]} = \sum_{j=1}^{\infty} \lambda_j \mathbb{Z}_{p,j}\mathbb{Z}_{p,j}' = \int_0^1 \int_0^1 Q_h^*(r, s)\, dW_p(r) dW_p'(s),$$

and $\mathbb{S}_{h,B}^{[p]} \xrightarrow{d} \mathbb{S}_{h,\infty}^{[p]}$, as $B \to \infty$, we can approximate the random quantity $\mathbb{S}_{h,\infty}^{[p]}$ using $\mathbb{S}_{h,B}^{[p]}$ for sufficiently large B. As a result, the critical values for the non-standard limits $\mathbb{W}_\infty$ and $\mathbb{T}_\infty$ can be easily generated by repeatedly drawing

$$\mathbb{Z}_p'[\mathbb{S}_{h,B}^{[p]}]^{-1}\mathbb{Z}_p \text{ and } \mathbb{Z}_p/\sqrt{\mathbb{S}_{h,B}^{[p]}},$$

respectively, where $\mathbb{Z}_p \sim N(0, I_p)$. Notably, the eigen decompositions in (15) is not required to implement fixed-smoothing asymptotic inference. Instead, computing (28) using random draws of standard normal vectors and the weight functions $Q_h(\cdot,\cdot)$ is sufficient to generate the non-standard fixed-smoothing limits. In (28), the number B represents the number of simulation draws used to approximate the Brownian motion process $W_p(\cdot)$ in $\mathbb{S}_{h,\infty}^{[p]}$. The choice B is related to the approximation of Brownian motion, and a larger B generally improves the quality of the approximation but increases computational burden. The literature on HAR inference suggests that setting B equal to T for moderately large T, say, $T \geq 200$, can simplify the computational burden while ensuring reliable finite-sample performance in fixed-smoothing asymptotic inference. See Sun (2014b) for details.

The result in Theorem 2-(b) implies that the Wald and t statistics using OS-LRV estimators converge to scaled versions of the standard F and t limits under fixed-smoothing (K) asymptotics. This result contrasts with Theorem 2-(a), which yields non-standard fixed-smoothing limits. To see this, note that the weight function $Q_h(r,s)$ for OS-LRV estimation indicates that $\lambda_j = 1/K$ for $j \in \{1,\ldots,K\}$ and $\lambda_j = 0$ for $j \geq K$. Consequently, we have

$$\mathbb{S}_{h,\infty}^{[p]} = \frac{1}{K}\sum_{j=1}^{K}\mathbb{Z}_{p,j}\mathbb{Z}_{p,j}' \overset{d}{=} \frac{1}{K}\mathbb{W}_p(K, I_p),$$

where $\mathbb{W}_p(K, I_d)$ is a scaled Wishart random matrix with K degrees of freedom. Therefore, the limit distribution $\mathbb{W}_\infty$ can be represented as $\mathbb{Z}_p'[\mathbb{W}_p(K, I_p)/K]^{-1}\mathbb{Z}_p$, which follows a scaled Hotelling's T-squared distribution. Using the well-known relationship between the T-squared and F, distributions up to the scale, we can express the fixed-smoothing limits for the Wald and t statistics in terms of the standard F and t distributions, respectively. Similar to the non-standard fixed-smoothing asymptotic critical values in Theorem 2-(a), using standard F and t critical values for OS-LRV is expected to reduce over-rejection when testing $H_0 : R\beta_0(\tau) = r$ in finite samples. This is because the F and t limits account for the estimation uncertainty of the nonparametric estimator $\hat{\Omega}(\tau)$ from the studentized HAR statistic. This feature contrasts with the conventional chi-square and normal approximations under increasing-smoothing asymptotics, which overlook the uncertainty arising from $\hat{\Omega}(\tau)$.

In summary, we establish that alternative fixed-smoothing asymptotic results and corresponding nuisance-parameter-free asymptotic inferences in the non-smooth QR setting. Importantly, we show that the Wald and t inferences using OS-LRV admit the exact F and t asymptotic critical values, which do not require any simulations for non-standard critical values in practice. While the kernel LRV yields non-standard fixed-smoothing critical values, practitioners can easily simulate them using only standard normal vectors and the kernel weight specified for $Q_h(\cdot,\cdot)$, which can provide more accurate Wald inferences than standard Chi-square tests in increasing-smoothing asymptotics.

4.2 Extension to quantile GMM and self-normalized inference

In this subsection, we discuss a possible extension of our HAR QR inference to non-smooth generalized method of moments (GMM) for quantile IV regression models. We also discuss the connection between our fixed-smoothing asymptotic inference and the self-normalization approach for QR.

Consider the non-smooth GMM for quantile models, e.g., de Castro et al. (2019):

$$\mathbb{E}[\tilde{Z}_t(\beta_0(\tau))] = \mathbb{E}[W_t(\tau - 1(y_t \le X_t'\beta_0(\tau))] = 0$$

if and only if $b = \beta_0(\tau)$ for a small neighborhood of $\beta_0(\tau)$. The covariate $X_t \in \mathbb{R}^d$ may include both endogenous and exogenous variables, and $W_t \in \mathbb{R}^m$ is a vector of exogenous instrumental variables that may include exogenous components in X_i. We assume that $m \ge d$, so that the QR-GMM model can be overidentified. The QR-GMM estimator minimizes a weighted quadratic norm of the sample moment vector:

$$\hat{\beta}(\tau) = \arg\min_{b\in\mathcal{B}} \left(\sum_{t=1}^{T} \tilde{Z}_t(b)\right)' \mathcal{M}_T \left(\sum_{t=1}^{T} \tilde{Z}_t(b)\right),$$

where $\tilde{Z}_t(b) = W_t(\tau - 1(y_t \le X_t'b)$, and $\mathcal{B} \subseteq \mathbb{R}^d$ is a compact parameter set. The $m \times m$ matrix $\mathcal{M}_T$ is a weight matrix that converges in probability to a strict positive definite matrix $\mathcal{M}$. Note that when $W_t = X_t$, the model reduces to the exactly identified QR model introduced in Section 2, where the use of $\mathcal{M}_T$ is unnecessary for obtaining $\hat{\beta}(\tau)$, as it can be computed by solving the sample moment condition in (1).

Under some regular conditions, the standard M-estimation theory with non-smooth moment function, e.g., Pakes and Pollard (1989), can be applied to obtain that

$$\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau)) = (G(\tau)'\mathcal{M}G(\tau))^{-1}G(\tau)'\mathcal{M}\underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t(\tau - 1(y_t \le X_t'\beta_0(\tau))}_{\xrightarrow{d} N(0,\tilde{\Omega}(\tau))} + o_p(1)$$

$$\xrightarrow{d} N(0, \tilde{\Sigma}(\tau)) \text{ with } \tilde{\Sigma}(\tau) = \tilde{D}(\tau)^{-1}\tilde{\Omega}(\tau)\tilde{D}(\tau)^{-1},$$

where

$$\tilde{D}(\tau) = (G(\tau)'\mathcal{M}G(\tau))^{-1}G(\tau)'\mathcal{M} \text{ and } G(\tau) = \mathbb{E}[f(0|W_t, X_t)W_tX_t'].$$

Let $\hat{\Omega}_h(\tau)$ denote the HAR LRV estimator of the long-run variance of the non-smooth moment process $\{\tilde{Z}_t(\beta_0(\tau)\}$, denoted as $\tilde{\Omega}(\tau)$. $\hat{\Omega}_h(\tau)$ can be obtained by replacing $\hat{Z}_t$ in (7) with $W_t(\tau - 1(y_t \leq X_t'\hat{\beta}(\tau))$. Also, we can formulate the Wald and t statistics for quantile-GMM by substituting $\hat{D}(\tau)$ with $(\hat{G}(\tau)'\mathcal{M}_T\hat{G}(\tau))^{-1}\hat{G}(\tau)\mathcal{M}_T$ in $\hat{\Sigma}(\tau) := \hat{D}(\tau)^{-1}\hat{\Omega}(\tau)\hat{D}(\tau)^{-1}$ from (26) and (27), respectively, where $\hat{G}(\tau)$ is a consistent estimate of the Jacobian matrix $G(\tau)$. The corresponding fixed-smoothing inference follows the same approach established in the previous subsection. One complication arises when estimating the feasible two-step efficient GMM, i.e., when $\mathcal{M}_T$ is formulated by HAR LRV estimates for the non-smooth moment process $\hat{\Omega}_h(\tau)$. In this case, the GMM weight matrix converges to a random matrix under fixed-smoothing asymptotics, which can lead to HAR inference results differing from those in the one-step GMM framework developed in this paper, e.g., Hwang and Sun (2017) and Hwang and Valdés (2023). Extensions in this direction are currently under development and will be addressed in a separate paper.

We next discuss the relationship between fixed-smoothing asymptotics and an alternative self-normalized (SN) approach in the time series literature, e.g., Shao (2010) and Zhou and Shao (2013). Instead of estimating the asymptotic variance of the QR estimator, the SN approach constructs a studentized statistic by utilizing recursive estimations of QR coefficients, starting from the subsample periods $\{1, \ldots, \lfloor cT \rfloor + 1\}$, where $c \in (0, 1)$ is a tuning parameter. Specifically, let $\{\hat{\beta}_{[s]}(\tau)\}_{s=\lfloor cT \rfloor+1}^{T}$ denote the sequence of QR estimators based on the subsamples of $\{1, \ldots, s\}$ for $s \leq T$. The SN statistic is then defined as

$$SN_T(\tau; c) := (R\hat{\beta}(\tau) - r)' \left(R\Upsilon(\tau; c)R'/T\right)^{-1} (R\hat{\beta}(\tau) - r), \tag{29}$$

where the self-normalized denominator $\Upsilon(\tau; c)$ is given by

$$\Upsilon(\tau; c) = \frac{1}{T^2} \sum_{s=\lfloor cT \rfloor+1}^{T} s^2 \left(\hat{\beta}_{[s]}(\tau) - \hat{\beta}(\tau)\right) \left(\hat{\beta}_{[s]}(\tau) - \hat{\beta}(\tau)\right)'. \tag{30}$$

A recent work by Hoga and Schulz (2025) establishes the asymptotic distribution of $SN_T(\tau; c)$ and shows that

$$SN_T(\tau; c) \xrightarrow{d} \mathbb{Z}_p'[\mathbb{SN}_{c,\infty}^{[p]}]^{-1}\mathbb{Z}_p,$$

where $\mathbb{Z}_p$ and $\mathbb{SN}_{c,\infty}^{[p]}$ are independent, and

$$\mathbb{SN}_{c,\infty}^{[p]} = \int_c^1 (W_p(s) - sW_p(1))(W_p(s) - sW_p(1))'ds.$$

To connect the SN approach to our fixed-smoothing asymptotic inference, we consider Wald statistics based on the kernel LRV, where the kernel is chosen as the Bartlett function, i.e., $Q_h(r, s) = k((r - s)/b)$ with $k(x) = 1(|x| \leq 1) \cdot (1 - |x|)$ and $h = 1/b$, with no truncation, meaning that the smoothing parameter h is chosen as 1 for $b = 1/h$. The corresponding HAR LRV can be formulated as

$$\hat{\Omega}_h(\tau) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h\left(\frac{t}{T}, \frac{s}{T}\right)\hat{Z}_t^c \hat{Z}_s^{c\prime} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(1 - \frac{|t-s|}{T}\right)\hat{Z}_t^c \hat{Z}_s^{c\prime} \tag{31}$$

$$= \underbrace{\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(1 - \frac{|t-s|}{T}\right) Z_t^c Z_s^{c\prime}}_{=\tilde{\Omega}_h(\tau)} + o_p(1), \tag{32}$$

where the last equation follows by Theorem 1. Using the same algebra shown in Kiefer and Vogelsang (2002), $\tilde{\Omega}_h(\tau)$ in (32) can be equivalently expressed as

$$\tilde{\Omega}_h(\tau) = \frac{2}{T}\sum_{s=1}^{T}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{s} Z_t^c\right]\left[\frac{1}{\sqrt{T}}\sum_{\tilde{t}=1}^{s} Z_{\tilde{t}}^c\right]',$$

which converges in distribution to the random limit:

$$\tilde{\Omega}_h(\tau) \xrightarrow{d} 2\Omega^{1/2}(\tau)\int_0^1 (W_p(s) - sW_p(1))(W_p(s) - sW_p(1))' ds \Omega^{1/2}(\tau)'.$$

This implies that the Wald statistic using (31) converges to the nonstandard limit:

$$\mathbb{W}_T(\tau) \xrightarrow{d} \mathbb{Z}_p'[\mathbb{S}_{h,\infty}^{[p]}]^{-1}\mathbb{Z}_p,$$

where $\mathbb{Z}_p$ and $\mathbb{S}_{h,\infty}^{[p]}$ are independent, and the random limit $\mathbb{S}_{h,\infty}^{[p]}$ defined in Theorem 2 can be represented as

$$\mathbb{S}_{h,\infty}^{[p]} = 2\int_0^1 (W_p(s) - sW_p(1))(W_p(s) - sW_p(1))' ds$$

As a result, up to the scaling factor 2, the random limit $\mathbb{S}_{h,\infty}^{[p]}$ derived in our fixed-smoothing asymptotic theory has the same form to that of $\mathbb{SN}_{c,\infty}^{[p]}$ with $c = 0$. The key difference is that the SN approach is infeasible when the tuning parameter c is set to 0, as $\Upsilon(\tau; 0)$ requires the estimation of $\hat{\beta}_{[s]}(\tau)$ starting from a single observation, i.e., $s = 1, 2, 3, \ldots T$, which is challenging to compute. Thus, in practice, selecting an appropriate nonzero tuning parameter $c \in (0, 1)$ is necessary to ensure that the recursive subsample estimators, $\hat{\beta}_{[\lfloor cT \rfloor+1]}(\tau), \hat{\beta}_{[\lfloor cT \rfloor+2]}(\tau), \ldots$, perform well in finite samples, and the implementation of the SN statistic in (29) is feasible. In contrast, our formulation of the HAR Wald statistic can encompass self-normalized asymptotic inference as a special case within QR-HAR inference by simply employing the Bartlett kernel function with $b = 1$ in HAR LRV estimation. Unlike the asymptotic inference based on the self-normalized statistic $SN_T(\tau; c)$, the fixed-smoothing asymptotic inference using (31) does not require c, which determines the initial subsample periods for the recursive estimation $\{\hat{\beta}_{[s]}(\tau)\}_{s=\lfloor cT \rfloor+1}^{T}$ in $\Upsilon(\tau; c)$.

5 Practical Implementation of QR-HAR Inference

In this section, we provide a practical recommendation for QR-HAR inferences, focusing on the choice of the smoothing parameter. Given the convenient features of standard F and t critical values and their improved finite-sample performance in the conditional mean regression setting, e.g., Sun (2013) and Lazarus et al. (2021), we focus on the asymptotic F and t QR-HAR inference presented in Theorem 2-(b), which can be implemented via OS-LRV.

The smoothing parameter h in the series OS-LRV estimator is equal to the number of basis functions K employed. The standard HAC approaches focus on the (asymptotic) mean squared error of the infeasible LRV estimator $\tilde{\Omega}_h(\tau)$ under increasing-smoothing asymptotics. These include Andrews (1991) and Phillips (2005) in the conditional mean regression, and Galvao and Yoon (2024) in the QR setting. Among these, we extend the OS-LRV result in Phillips (2005) to our QR setting as follows: Assume that T and K increase to infinity such that $K/T \to 0$, then we have that

$$\begin{aligned} MSE(\tilde{\Omega}_h(\tau)) &= \mathbb{E}\left[vec(\tilde{\Omega}_h(\tau) - \Omega(\tau))'\mathcal{W}vec(\tilde{\Omega}_h(\tau) - \Omega(\tau))\right] \\ &= \frac{K^4}{T^4}(vec(B(\tau))'\mathcal{W}vec(B(\tau))) + \frac{1}{K}\mathrm{Tr}[\mathcal{W}(I_{d^2} + \mathbb{K}_{dd})(\Omega(\tau) \otimes \Omega(\tau))] \\ &\quad + o\left(\frac{K^4}{T^4} + \frac{1}{K}\right), \end{aligned} \tag{33}$$

where $\mathcal{W}$ is a $d^2 \times d^2$ weight matrix chosen by practitioners, $\mathbb{K}_{dd}$ is the $d^2 \times d^2$ commutation matrix, $vec(\cdot)$ denotes the column-by-column vectorization operator. The $d \times d$ matrix $B(\tau)$ captures the bias of $\tilde{\Omega}_h(\tau)$, which is of order K^2/T^2:

$$B(\tau) := \lim_{T\to\infty} \left(\frac{T}{K}\right)^2 \left(\mathbb{E}[\tilde{\Omega}_h(\tau)] - \Omega(\tau)\right) = -\frac{\pi^2}{6} \sum_{j=-\infty}^{\infty} j^2 \Gamma_j(\tau).$$

It is then straightforward to show that the leading term of (33) is minimized at:

$$K^*_{MSE} = \left\lceil \left(\frac{\mathrm{Tr}[\mathcal{W}(I_{d^2} + \mathbb{K}_{dd})(\Omega(\tau) \otimes \Omega(\tau))]}{4vec(B)'\mathcal{W}vec(B(\tau))}\right)^{1/5} \cdot T^{4/5} \right\rceil.$$

The optimal choice K^*_{MSE} that minimizes asymptotic MSE is motivated by classical work on spectral density estimation, such as Parzen (1957) and Hannan (1970), where their main focus is on the spectral density of $\{Z_t\}$ at the zero point, which is equal to the LRV $\Omega(\tau)/2\pi$. However, this choice may not be ideal for hypothesis testing and constructing confidence intervals, which are the ultimate goals in robust QR inference. In the conditional mean regression setting, it has been shown that HAC-type inferences using the MSE-optimal smoothing parameter can have large size distortions in finite samples, e.g., Lazarus et al. (2019).

An alternative to the MSE-based approach is to select the smoothing parameter K that minimizes testing-oriented loss functions. The construction of the testing-oriented smoothing parameter begins by transforming the original QR inference into the following p-dimensional location model:

$$W_t = \mu_0 + v_t \in \mathbb{R}^p \text{ for } t \in \{1, \dots, T\} \tag{34}$$

with $\mu_0 = \mathbb{E}[W_t]$. We are interested in testing H_0: $\mu_0 = 0$ against H_1: $\mu_0 \neq 0$, with corresponding Wald statistic defined as:

$$\begin{aligned} \tilde{\mathbb{W}}_T(\tau) &= \left(\sqrt{T}(\bar{W}_T - \mu_0)\right)' \tilde{\Sigma}^{-1}(\tau) \left(\sqrt{T}(\bar{W}_T - \mu_0)\right) \\ &= \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_t\right)' \tilde{\Sigma}^{-1}(\tau) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_t\right), \end{aligned} \tag{35}$$

where $\tilde{\Sigma}(\tau) = K^{-1} \sum_{j=1}^{K} \tilde{U}_j \tilde{U}_j'$ for $\tilde{U}_j = T^{-1/2} \sum_{t=1}^{T} \Phi_j(t/T)\,(v_t - \bar{v}_T)$.

The error process $\{v_t\}$, defined on the same probability space as the original data $\{Y_t, X_t'\}$, has zero mean and is assumed to be a covariance-stationary process that shares the same autocovariance structure as $\{R\,(D(\tau))^{-1}\,Z_t\}$. This implies that

$$\begin{aligned} \sqrt{n}(\bar{W}_T - \mu_0) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_t \xrightarrow{d} N(0, R\Sigma(\tau)R'); \\ \left(\frac{K-p+1}{K}\right) &\tilde{\mathbb{W}}_T(\tau) \xrightarrow{d} p\mathcal{F}_{K-p+1}, \end{aligned}$$

under $H_0 : \mu_0 = 0$. Therefore, the testing problem in (34) is asymptotically equivalent to testing the original null hypothesis in QR, $H_0 : R\beta_0(\tau) = r$. Also, the Wald statistic $\tilde{\mathbb{W}}_T(\tau)$ shares the same weak convergence limit as $\mathbb{W}_T(\tau)$ under fixed-smoothing asymptotics.

We consider the following set for the sequence of local data generation:

$$\mathcal{C}_T(\delta^2) := \left\{ \frac{(R\Sigma(\tau)R')^{1/2} u}{\sqrt{T}}; u \in \mathbb{R}^p \right\}, \tag{36}$$

where u is uniformly distributed on a p-dimensional sphere with radius $\delta \geq 0$. We will choose K to follow the classical Neyman-Pearson (NP) principle that maximizes a power of HAR inference subject to a level-α of the test. Specifically, given $\delta > 0$, the optimal K in HAR-inference satisfies:

$$\text{Maximize } P(\text{Reject } H_0 | \text{When } H_1(\delta^2) : \mu_0 \in \mathcal{C}_T(\delta^2) \text{ with } \delta > 0 \text{ is true}) \tag{37}$$

$$\text{subject to } P(\text{Reject } H_0 | \text{When } H_0 : \mu_0 = 0 \text{ is true}) \leq \alpha. \tag{38}$$

This genuine notion of the testing-optimal K was first introduced by Sun et al. (2008) in the conditional mean regression setting with kernel-HAR inference and was further extended to the OS-HAR by Sun (2011, 2013). See also Lazarus et al. (2018) and Lazarus et al. (2021) for other versions of the testing-optimal criterion within a non-NP framework. We will show that the NP-approach can be adapted to our HAR-QR-setting.

We first want to approximate the exact probabilities in (37) and (38) using higher-order expansion of them. To achieve this goal, we assume that $\{v_t\}$ is a Gaussian process. This assumption for $\{v_t\}$ is a strong restriction, but it should not be interpreted as applying to the true distribution for its approximating target $\{R\,(D(\tau))^{-1}\,Z_t\}$ in QR. This is because our fixed-smoothing asymptotic critical value $p\mathcal{F}^{\alpha}_{p,K-p+1}$ is valid for a broader range of non-Gaussian data. Additionally, using the infeasible quantity $D(\tau)$ in $\{R\,(D(\tau))^{-1}\,Z_t\}$ isolates the problem of LRV estimation in HAR-QR inference from the non-parametric estimation quantity $\hat{D}(\tau)$. Under these assumptions, we can examine the higher-order expansion of HAR-inference's null rejection and power probabilities. This enables us to derive a closed-form formula for the Neyman-Pearson problem in (37) and (38) without relying on complex Edgeworth-type expansions. We could establish the higher-order expansion without these assumptions, but it would require including many additional terms that are irrelevant to the smoothing parameter in the expansion of the probability distribution for $\tilde{\mathbb{W}}_T(\tau)$, as shown in Sun and Phillips (2008).

Let $G_{p,\delta^2}(\cdot)$ be the cumulative distribution function (CDF) of $\chi^2_p(\delta^2)$, which is the non-central chi-squared distribution with the non-centrality parameter δ^2. When $\delta^2 = 0$, $G_p(\cdot) := G_{p,\delta^2}(\cdot)$ denotes the cumulative distribution function of a χ^2-random variable with p-degrees of freedom, with the corresponding $(1-\alpha)$-quantile χ^{α}_p. Then, the result from Sun (2011, Theorem 5) can be applied to our setting in (34) as follows: If T and K increase to infinity such that $K/T \to 0$ under the data generation process of $\mu_0 \in \mathcal{C}_T(\delta^2)$, we have that

$$P\left(\left(\frac{K-p+1}{K}\right)\tilde{\mathbb{W}}_T(\tau) > z\right) = 1 - G_{\delta^2,p}(z) - \left(\frac{K}{T}\right)^2 \left(G'_{\delta^2,p}(z)z\right)\bar{B}(\tau) - \frac{1}{K}\left(G''_{\delta^2,p}(z)z^2\right) + o\left(\frac{K^2}{T^2} + \frac{1}{K}\right) + O\left(\frac{1}{T}\right), \tag{39}$$

where $G'_{\delta^2,p}(\cdot)$ and $G''_{\delta^2,p}(\cdot)$ denote the first and second derivatives of $G'_{\delta^2,p}(\cdot)$, respectively. The constant $\bar{B}(\tau)$ is defined as follows:

$$\bar{B}(\tau) := \bar{B}(R, D(\tau), \Omega(\tau), B(\tau)) = \frac{1}{p}\cdot \mathrm{Tr}([(R\tilde{B}R')][R\Sigma(\tau)R']^{-1}) \tag{40}$$

with $\tilde{B} := (D(\tau))^{-1}\,B\,(D(\tau))^{-1}$. In (39), the second and the third terms on right-hand side captures two different effects:

$$\left(\frac{K}{T}\right)^2 G'_{\delta^2,p}(z)z\bar{B}(\tau): \text{ Bias effect of } R\tilde{\Sigma}(\tau)R';$$
$$\frac{1}{K}G''_{\delta^2,p}(z)z^2: \text{ Variance effect of } R\tilde{\Sigma}(\tau)R',$$

which implies that there is an opportunity to select K to balance these effects under the testing-oriented loss function specified in (37) and (38). Now, we shall go one step further by considering the asymptotically level-α HAR inference that uses the fixed-K critical value. Let $\mathcal{F}^{1-\alpha}_{p,K-p+1}$ be $(1-\alpha)$-quantile of $\mathcal{F}_{p,K-p+1}$. Using the relation shown in equation (5) of Sun (2013), we can show that, if $\mu_0 = 0$,

$$P\left(\left(\frac{K-p+1}{K}\right)\tilde{\mathbb{W}}_T(\tau) > p\mathcal{F}^{\alpha}_{p,K-p+1}\right) = \alpha - \left(\frac{K}{T}\right)^2 \left(G'_p(\chi^{\alpha}_p)\chi^{\alpha}_p\right)\bar{B}(\tau) + o\left(\frac{K^2}{T^2} + \frac{1}{K}\right) + O\left(\frac{1}{T}\right), \tag{41}$$

as $K \to \infty$, where χ_p^α is $(1-\alpha)$-quantile of χ_p^2 random variable. Thus, using the fixed-K critical value can remove the variance effect under H_0, which indicates that the fixed-K critical value is second-order correct. Also, we can obtain the second order approximation of the null rejection probability in (38) for HAR-inference:

$$e_I(K) := \alpha - \left(\frac{K}{T}\right)^2 \left(G_p'(\chi_p^\alpha)\chi_p^\alpha\right) \bar{B}(\tau).$$

To approximate the power of HAR inference, under $H_1(\delta^2)$ in (36), it can be shown that

$$\begin{aligned} P\left(\left(\frac{K-p+1}{K}\right)\tilde{\mathbb{W}}_T(\tau) > p\mathcal{F}_{p,K-p+1}^\alpha\right) &= 1 - G_{p,\delta^2}\left(\chi_p^\alpha\right) - \left(\frac{K}{T}\right)^2 \left(G_{p,\delta^2}'(\chi_p^\alpha)\chi_p^\alpha\right) \bar{B}(\tau) \\ &\quad - \frac{(\chi_p^\alpha)^2}{K} Q(p,\delta^2) + o\left(\frac{K^2}{T^2} + \frac{1}{K}\right) + O\left(\frac{1}{n}\right), \end{aligned}$$

where $Q_{p,\delta^2}(\chi_p^\alpha) = G_{p,\delta^2}''(\chi_p^\alpha) - (G_p''(\chi_p^\alpha)/G_p'(\chi_p^\alpha))G_{p,\delta^2}'(\chi_p^\alpha)$. Note that the expansions exactly coincide with (41) under the null when $\delta = 0$. Using the relationship $Q(p,\delta^2) = (\delta^2/2)(G_{(p+2),\delta^2}'(\chi_p^\alpha)/\chi_p^\alpha)$, as shown in Sun (2011), the result leads us to approximate the power probability in (37):

$$\begin{aligned} &1 - \text{Type II error} \simeq 1 - e_{II}(\delta^2, K); \\ &e_{II}(\delta^2, K) := G_{p,\delta^2}\left(\chi_p^\alpha\right) + \left(\frac{K}{T}\right)^2 \left(G_{p,\delta^2}'(\chi_p^\alpha)\chi_p^\alpha\right) \bar{B}(\tau) + \frac{\delta^2}{2K}\left(G_{(p+2),\delta^2}'(\chi_p^\alpha)\chi_p^\alpha\right). \end{aligned}$$

The approximated quantities for Type I and Type II errors directly represent size and power probabilities of HAR-inference. Also, they allow us to provide a feasible solution of the optimal testing problem in (37) and (38). Let κ denote a user-chosen tuning parameter value such that $\kappa > 1$. We then solve the constrained minimization problem, which takes the same form as in Sun (2011):

$$\tilde{K}^* = \arg\min_K e_{II}(\delta^2, K) \text{ such that } e_I(K) \leq \kappa\alpha.$$

Here, introducing the tuning parameter κ reflects that the approximated probability can have some error to satisfy the NP-principle's nominal level constraint in (38). The constrained minimization problem yields the following explicit form of the solution:

$$\tilde{K}^* = \begin{cases} \left[\frac{\delta^2 G_{(p+2),\delta^2}'(\chi_p^\alpha)}{4G_{p,\delta^2}'(\chi_p^\alpha)|\bar{B}(\tau)|}\right]^{\frac{1}{3}} \cdot T^{\frac{2}{3}}, \text{ if } \bar{B}(\tau) > 0; \\ \left[\frac{(\kappa-1)\alpha}{G_p'(\chi_p^\alpha)\chi_p^\alpha|\bar{B}(\tau)|}\right]^{\frac{1}{2}} \cdot T, \quad \text{ if } \bar{B}(\tau) < 0 \end{cases}. \tag{42}$$

and, the testing optimal K^* is formulated as

$$K^* = \max\left\{\left\lceil \tilde{K}^* \right\rceil, \; p+4\right\},$$

so that the asymptotic distribution $p\mathcal{F}_{p,K^*-p+1}$ has a finite first moment. The practical application of the optimal K^*, a requires values for several parameters, including δ^2, π, and $\bar{B}$. Readers are referred to subsection 8.1 of the Appendix, where we provide a data-driven formula for these values.

6 Monte Carlo Evidence

In this section, we perform Monte Carlo simulations to validate our theoretical findings and confirm the favorable finite-sample performance of the QR-HAR inference method.

6.1 Comparison with HAR conditional mean regression

We consider the following DGP, considered in Sun (2014a):

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \epsilon_t \text{ for } t \in \{1, \ldots, T\}, \tag{43}$$

where $x_{j,t}$ and ϵ_t follow AR(1) processes,

$$x_{j,t} = \rho_x x_{j,t-1} + e_{j,t} \text{ and } \epsilon_t = \rho \epsilon_{t-1} + e_{0,t} \tag{44}$$

for $j \in \{1, 2, 3\}$, $\rho_x = 0.75$, $\rho \in \{0.25, 0.50, 0.75, 0.90\}$, and $(e_{1,t}, ..., e_{d,t})' \overset{\text{i.i.d.}}{\sim} N(0, \sqrt{1-\rho_x^2} \cdot I_d)$. For the innovation term $e_{0,t} \overset{\text{i.i.d.}}{\sim} F_e$, we choose its probability distribution F_e to have a zero mean and unit variance, with distributions specified as:

$$F_e \sim N(0,1) \text{ or } F_e \sim \frac{\chi_2(1) - 1}{\sqrt{2}} \text{ or } F_e \sim \frac{\mathcal{T}_3}{\sqrt{3}}. \tag{45}$$

The initial value ϵ_0 is drawn from its unconditional distribution, e.g., $N(0, 1/(1-\rho^2))$ with $F_e \sim N(0,1)$.The unknown parameter vector is $\theta = (\alpha, \beta_1, \beta_2, \beta_3)' \in \mathbb{R}^d$ with $d = 4$, and is set to be $(0, 1, 1, 1)'$. Since the conditionally homoskedastic e_t is independent of $x_{j,t}$, the true parameter for the QR coefficient vector, $(\beta_1(\tau), \beta_2(\tau), \beta_3(\tau))'$, is equal to $(1, 1, 1)'$ for all quantile levels $\tau \in (0, 1)$. The null hypotheses of interest are

$$H_{01} : \beta_1(\tau) = 1, \tag{46}$$

$$H_{02} : \beta_1(\tau) = \beta_2(\tau) = 1, \tag{47}$$

$$H_{03} : \beta_1(\tau) = \beta_2(\tau) = \beta_3(\tau) = 1, \tag{48}$$

where the numbers of joint hypotheses are $p = 1$, 2, and 3, respectively. When $d = 2$ and $F_v \sim N(0, (1-\rho^2))$, our simulation designs in (43) and (44) are equivalent to the DGP in Galvao and Yoon (2024), which examined the performance of the QR-HAC t statistic using standard normal critical value. Our designs are also comparable to the AR(2) designs in Gregory et al. (2018) which employ the same regression error specifications outlined in (45). Our simulation designs, including hypothesis testing (46)–(48), encompass multivariate Wald inference for both QR-HAC and QR-HAR approaches.

We examine the empirical rejection probability (ERP) of four types of Wald inferences for testing (46)–(48), with the nominal significance level α set at 5%. The first Wald inference, denoted as `HAR-M`, utilizes the existing HAR-Wald inference method, which employs ordinary least square (OLS) estimates with OS LRV estimates and t and F critical values, as described in Müller (2007) and Sun (2013). It applies the smoothing parameter K^*, determined by minimizing the testing-oriented selection rule, as outlined in Ye and Sun (2018). The next two procedures, denoted as `QR-HAC` and `QR-HAR`, are based on QR estimates at $\tau = 0.50$, (i.e., median regression) and implement HAC- and HAR-based QR-Wald inference, respectively. Specifically, `QR-HAC` employs kernel LRV estimates using the Bartlett kernel and chi-square critical values derived from conventional increasing-smoothing (or small-b) asymptotics. The smoothing parameter b^* is selected by minimizing the asymptotic mean squared error (AMSE) using the optimal formula provided in Galvao and Yoon (2024). `QR-HAR` is based on the OS-LRV estimate and the fixed-smoothing asymptotic result in Theorem 2-(c), which uses (scaled) F-critical values. The smoothing parameter K^* is obtained using the testing-oriented rule with the data-dependent formulation specified in the previous section.

The final method, denoted as `SETBB`, employs the smooth extended tapered block bootstrap approach developed by Gregory et al. (2018). It is designed to improve inference in time series QR by extending the tapered moving block bootstrap method, incorporating smoothing for both data blocks and individual observations. Implementing `SETBB` requires two smoothing parameters: one for individual data and one for block length. For the smoothing parameter for individual data, we use the formula recommended in Section 4 of Gregory et al. (2018). For the block length, we set it as b^*T, where b^* is the AMSE-optimal smoothing parameter in `QR-HAC`. We consider sample size $T \in \{200, 400, 800\}$, with $10,000$ replications in all Monte Carlo simulations. The results are provided in Tables 1–3 for the null hypotheses (46)–(48), with $p \in \{1, 2, 3\}$ and error distributions specified in (45).

Our results first indicate that Wald inferences for both conditional mean regression (OLS) and QR perform well when the degree of time series correlation ρ is low, i.e., $\rho = 0.25$, as their ERPs are close to the nominal level 5%. For example, Table 1 for normal errors and $\rho = 0.25$ shows that ERPs for `HAR-M`, `QR-HAC`, and `QR-HAR` with $T = 200$ and $p = 1$ are 7.2%, 8.1%, and 7%, respectively. The size distortions reduce to 6.9%, 7.1%, and 6.3%, as the sample size increase to $T = 400$. Not surprisingly, the ERPs for OLS-based `HAR-M` increase when the errors are non-Gaussian, while the ERPs for QR-based inference are less sensitive. For instance, results in Table 1–3 show that the ERP for HAR-M with $p = 3$ and $\rho = 0.25$ increases from 8.6% to 10.3% and 10.6% when the error distributions shift to chi-square and $\mathcal{T}_3$, respectively. On the other hand, our results show that the ERPs for QR-HAR inference remain steady, ranging from 8.9% to 9.8% across all error specifications.

Tables 1–3 also indicate that the size distortions of OLS-based Wald inference, `HAR-M`, substantially increase, as ρ increase. Specifically, Table 1 for normal errors shows that the ERPs for `HAR-M` with $T = 200$ and $p = 1$ increases from 7.2% to 14.1% as ρ increases from 0.250 to 0.50, and further increase to 56.3%

when ρ becomes 0.90. One reason for this is that the variance of innovation, $var(\epsilon_t) = 1/(1-\rho^2)$ approaches to infinity as ρ increases to one. Thus, the time series correlation problem in the HAR conditional mean regression also affects the larger (unbounded) second moment of the underlying moment process in our DGPs. In contrast, our results show that QR-based Wald inference have substantially lower size distortions than `HAR-M` when ρ increases. For instance, ERPs of `QR-HAC` are around 9.5%–13.4% and those for `QR-HAR` are around 7.3%–9.2% for $T = 200$, $\rho \in \{0.50, 0.75, 0.90\}$, and $p = 1$ in Table 1. Also, Tables 1–3 indicate that, for a given ρ value, the ERPs of `QR-HAC` and `QR-HAR` approach the nominal level of 5%, as sample size grows.

While QR-based Wald inference substantially reduces the finite-sample size distortions observed in OLS-based Wald inference, our results indicate that the HAC-based QR Wald inference, `QR-HAC`, can still suffer from severe size distortions when the degree of temporal dependence is high, e.g., $\rho = 0.90$. We also observe that size distortions become more serious as the number of testing parameters p increases. A key reason for the failure of `QR-HAC` is that the non-parametric LRV estimate exhibits high variation in finite samples, but these variations are not accounted for in the chi-square critical value used for `QR-HAC`. This result naturally motivates the implementation of the HAR-based QR Wald inference, `QR-HAR`, where its OS-LRV estimate utilizes standard (scaled) F critical values and a testing-oriented smoothing parameter choice, rather than the AMSE-optimal rule.

In Tables 1–3, we find that the size distortions of `QR-HAC` are reduced in `QR-HAR`, especially for sample sizes T ranging 200 and 400. For example, Table 1 with $p = 1$ and $T = 200$ reports that the ERPs of `QR-HAR` reduce those of `QR-HAC` from 8.1%–9.5% to 7.0%–7.3% at $\rho \in \{0.25, 0.50\}$, and 11.9%–13.4% to 8.9–9.2% at $\rho \in \{0.75, 0.90\}$. The reductions in size distortions become more pronounced as ρ increases, as shown in Table 1, as well as Tables 2 and 3 for other DGPs. Additionally, results with different values of p show that finite-sample improvements also increase when the number of testing hypotheses p increases, consistent with existing HAR literature, such as Sun (2011, 2013) and Hwang and Sun (2018).

In summary, the HAR-QR approach significantly reduces the empirical size distortions of the HAR conditional mean regression approach. Its finite-sample performance remains robust to the time series persistence of the regression error, even when the error distribution is non-Gaussian, asymmetric, or contains outliers. The HAR-based QR Wald inference not only addresses serial correlation but also mitigates the effects of heavy-tailed distributions (unbounded second moments), which negatively affect the performance of the HAR conditional mean regression. Additionally, our QR-HAR approach, based on the asymptotic F-test with a data-driven, testing-optimal smoothing parameter, performs well in finite samples, reducing the ERPs of the conventional chi-square test in the QR-HAC approach across various DGPs.

Lastly, we report the results of the bootstrap approach in Gregory et al. (2018), `SETBB`, as presented in Tables 1–3. The results indicate that `SETBB` can serve as an alternative to HAR inference, as it yields conservative finite-sample sizes when the degree of serial correlation is moderate. For example, Table 1 with $p = 1$ and $T = 200$ shows that the ERPs of `SETBB` range from 2.0% to 4.0% for $\rho \in \{0.25,\ 0.50\}$. However, our findings indicate that the size distortions of `SETBB` are larger than those of the QR-HAR approach when the degree of serial correlation is high, e.g., $\rho \in \{0.75,\ 0.90\}$, with ERPs ranging from 9.1% to 17.9% with $p = 1$ and $T = 200$. These numerical findings suggest that QR-HAR Wald inference, `QR-HAR`, is comparable to the bootstrap approach in Gregory et al. (2018), `SETBB`, when serial correlation is moderate and outperforms the bootstrap approach when serial correlation is strong. Importantly, our approach employs studentized Wald inference, whereas `SETBB` is limited to unstudentized statistics, as its bootstrap critical values are designed for unstudentized statistics.

6.2 Performance under heteroskedastic error variances

Our next simulation designs considers the same DGPs for the regressors $(x_{1,t}, x_{2,t}, x_{3,t})'$ and errors ϵ_t as in (44), but replaces (43) with

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + (\eta_0 + \eta_1 x_{1,t})\epsilon_t \text{ for } t \in \{1, \ldots, T\}, \tag{49}$$

where $\eta_0 = 2$ and $\eta_1 = 0.5$. The innovation term $e_{0,t}$ from (44) follows the distribution $e_{0,t} \sim N(0, \sqrt{1-\rho_e^2})$. In this case, the QR coefficient vector is given by $(\beta_1(\tau), \beta_2(\tau), \beta_3(\tau))' = (\beta_1 + \eta_1 \Phi^{-1}(\tau), \beta_2, \beta_3)'$, where $\Phi(\cdot)$ is the CDF of standard normal random variable, so the τ-QR coefficient for $x_{1,t}$ differs from that of the conditional mean regression when $\tau \neq 0.5$. The null hypotheses of interest are

$$H_{01} : \beta_1(\tau) = \beta_1 + \eta_1 \Phi^{-1}(\tau), \tag{50}$$

$$H_{02} : \beta_1(\tau) = \beta_1 + \eta_1 \Phi^{-1}(\tau) \text{ and } \beta_2(\tau) = \beta_2, \tag{51}$$

for $p = 1$, 2, respectively, and $(\beta_1, \beta_2, \beta_3)' = (1, 1, 1)'$.

Similar to the previous subsection, we examine the empirical rejection probability (ERP) of inference using `QR-HAC`, `QR-HAR`, and `SETBB`, but exclude `HAR-M` as the inference cannot test the heterogeneous quantile effects in (50) and (51) for $\tau \neq 0.50$. The results are reported in Tables 4 and 5. They indicate that the HAC inference in QR, `QR-HAC`, can achieve ERPs close to the nominal level for moderate levels of serial correlation with a large sample size, e.g., $\rho \in \{0.25, 0.50\}$ and $T = 800$, when testing a single hypothesis ($p = 1$). However, there are significant size distortions for `QR-HAC`, especially with moderate sample sizes and a higher degree of persistence, such as $T = 200$ with $\rho \in \{0.75, 0.90\}$. The HAR inference in QR, `QR-HAR`, can reduce the size distortions observed in `QR-HAC` for these cases. For example, when $\rho \in \{0.75, 0.90\}$ and $n = 400$ with $\tau = 0.75$, Table 4 that the ERPs for QR-HAC, ranging from 11.4% to

15.2%, are reduced to 9.4%–10.4% by QR-HAR. Similar to the previous analysis with $\tau = 0.50$, Tables 4 and 5 indicate that the extent of finite-sample improvements increases when multiple hypotheses are tested in Wald inference. The performance of the bootstrap approach, `SETBB`, exhibits similar findings as in the previous section. Its ERPs are most accurate in size when there is a moderate degree of serial dependence, such as $\rho \in \{0.75, 0.90\}$. However, `SETBB` shows more size distortions than `QR-HAR` when $\rho = 0.90$ and $\tau = 0.75$, as shown in Table4.

Finally, we note that both Wald inference methods, QR-HAC and QR-HAR, are subject to size distortions when the quantile level τ focuses on the tails, such as $\tau = 0.90$. However, the magnitude of these distortions decreases as the sample size increases. We also observe that the bootstrap approach, `SETBB`, is less sensitive when making QR inference at the tails and provides more accurate sizes than QR-Wald inference. In contrast to `SETBB`, the primary source of size distortions in QR-Wald-inference arises from finite-sample variability in the nonparametric estimation of the QR Hessian matrix in (52). Employing resampling methods in QR can provide more accurate estimates of asymptotic variance, thereby enhancing the QR-HAR Wald inference proposed in this paper. We leave this for future research.

7 Conclusion

In this paper, we establish an alternative fixed-smoothing asymptotic theory for quantile regression (QR) in time series under unknown form of weak temporal dependence. We show that the long-run variance (LRV) estimate in QR weakly converges to a random matrix scaled by the true LRV, and the corresponding QR-Wald statistics weakly converge to non-standard limits. Building on this result, we extend HAR inference in the conditional mean regression models provided by Sun (2014a & b) and Lazarus et al. (2021) to QR setting. Additionally, we show that the Wald and t inferences using orthonormal series LRV (OS-LRV) admit the standard F and t asymptotic critical values, which do not require any simulations for non-standard critical values in practice.

Regarding the choice of the smoothing parameter in OS-LRV, we develop an optimal smoothing selection rule that addresses the central concern of hypothesis testing, following the classical Neyman-Pearson principle. Based on approximated Type I and Type II errors, we provide a closed-form formula for the testing-optimal smoothing parameter.

Our simulations show that the asymptotic F-test, with a data-driven, testing-optimal smoothing parameter, performs well in finite samples, reducing the size distortions of both OLS-based HAR Wald inference and HAC Wald inference in QR across various data generation processes. The finite-sample improvements of our QR-HAR approach are more pronounced when the sample size is moderate, the degree of temporal dependence increases, and the number of null hypothesis parameters grows. Given its convenience and improved finite-sample performance, we recommend using OS-LRV in HAR inference, along with our data-driven smoothing parameter selection rule, for the QR setting.

8 Appendix of tables, formulas, and proofs

8.1 Data-driven implementation of QR-HAR inference

In the formula for the testing-oriented smoothing parameter (42), the function $G'_{p,\delta^2}(\cdot)$ is the pdf of the non-central χ^2-random variable with degrees of freedom p, and the noncentrality parameter δ^2. It can be written as

$$G'_{p,\delta^2}(x) = \frac{1}{2}\exp\left(-\frac{(x+\delta^2)}{2}\right)\cdot\left(\frac{x}{\delta^2}\right)^{\frac{p-2}{4}} I_{(p/2-1)}\left(\sqrt{\delta^2 x}\right),$$

where $I_\nu(s) = (s/2)^\nu \sum_{j=0}^{\infty}(s^2/4)^j/(j!\Gamma(\nu+j+1))$ with $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ is a modified Bessel function of the first kind. Given χ_p^α, the value of $G'_{p,\delta^2}(\chi_p^\alpha)$ can be computed using popular programming software such as `MATLAB` and `R`, which have functions `ncx2pdf` and `dchisq` for this purpose, respectively. Using (42). For the tuning parameter δ^2, according to its definition in (36), the value of δ^2 indicates the degree to which the true parameter $R\beta_0(\tau)$ deviates from the original null hypothesis $H_0 : R\beta_0(\tau) = r$ under the following alternative hypothesis:

$$H_1(\delta^2) : R\beta_0(\tau) = r + \frac{(R\Sigma(\tau)R')^{1/2}u}{\sqrt{T}},$$

where $||u||^2 = \delta^2$. In principle, δ^2 can be chosen to reflect the scientific interest or economic significance implied by the hypothesis test. When the information is not available, however, we can set a value to satisfy the rule $P(\chi_p(\delta^2) > \chi_p^{1-\alpha}) = 0.75$, as suggested by Sun (2011, 2014). Note that this value of δ^2 corresponds to the point at which the infeasible χ^2-test for QR, using the known asymptotic covariance matrix $R\Sigma(\tau)R'$, achieves 75% of the local asymptotic power.

For the parameter κ, which is greater than one, the value reflects the tolerance level for the deviation of the second-order approximated type I error probability $e_I(K)$ from the nominal level α. For example, if $\kappa = 1.10$ and the nominal level is 5% (i.e., $\alpha = 0.05$), the optimal HAR inference aims to control the type I error such that it does not exceed 5.5%. We may allow κ to depend on the sample size n. For larger sample sizes, e.g., T is more than 500, κ can be set to smaller values such as 1.05. We select a smaller value of κ than what is recommended for the conditional mean regression case, e.g., Ye and Sun (2013) with $\kappa = 1.15$, because the actual QR-HAR statistic is expected to have more approximation errors, leading to a greater deviation from the Gaussian location benchmark model.

The feasible estimation of the optimal smoothing parameter K^* requires an estimation of the $p\times p$ matrix $\bar{B}$ specified in (40), which includes estimations of $D(\tau), \Omega(\tau)$, and $B(\tau)$. For $\hat{D}(\tau)$, the consistent Powell estimator takes the form $D(\tau)$:

$$\hat{D}(\tau) = \frac{1}{Tl_T}\sum_{t=1}^{T} K\left(\frac{y_t - X_t'\hat{\beta}(\tau)}{l_T}\right) X_t X_t', \tag{52}$$

where $K(u) := 2^{-1}1(|u| \leq 1)$ is the uniform kernel. Regarding the choice of bandwidth parameter l_T, our Monte Carlo simulations indicate that the bandwidth rules proposed by Hall and Sheather (1988) and Bofinger (1975) perform well even in the presence of dependent errors. This aligns with the findings of Galvao and Yoon (2024) in their investigation of the QR-HAC approach. The formulas for the Hall-Sheather and Bofinger rules, as provided in Koenker (2005), can be specified in the Gaussian case as follows:

$$\hat{l}_T = T^{-1/5} \cdot \left\{ \frac{4.5\psi^4(\Phi^{-1}(\tau))}{(2\Phi^{-1}(\tau)^2+1)^2} \right\}^{1/5};$$
$$\hat{l}_T = T^{-1/3} \cdot \left\{ z_\alpha^{2/3} \left(\frac{1.5\psi^2(\tau)}{2(\psi'(\tau)/\psi(\tau))^2 + (\psi'(\tau)/\psi(\tau) - \psi''(\tau)/\psi(\tau))} \right)^{1/3} \right\},$$

respectively, where $\psi(\cdot)$ is the standard normal probability density function, and z_α denotes the $(1-\alpha)$ quantile for a standard normal random variable. These data-dependent estimates of $\hat{D}(\tau)$ can be easily executed using the R package `quantreg` and were implemented in our Monte Carlo simulations.

For values of $\Omega(\tau)$ and $B(\tau)$, Phillips (2005), Sun (2013), and Hwang and Valdés (2022) propose using VAR(1) approximation of the regression score process, i.e., $Z_t = A_z Z_{t-1} + w_{zt}$ for $A \in \mathbb{R}^{d\times d}$ and $\mathbb{E}[w_{zt}w_{zt}'] = \Upsilon$, and utilize their implied parametric forms given below:

$$\Omega_{\text{var}}(\tau) = (I_d - A_z)^{-1}\Upsilon(I_d - A_z')^{-1}; \tag{53}$$
$$B_{\text{var}}(\tau) = -\frac{\pi^2}{6}(I_d - A_z)^{-3}\left(A_z\Upsilon + A_z^2\Upsilon A_z' + A_z^2\Upsilon - 6A_z\Upsilon A_z' \right. \tag{54}$$
$$\left. +\Upsilon(A_z')^2 + A_z\Upsilon(A_z')^2 + \Upsilon A_z'\right)(I_d - A_z')^{-3}.$$

The VAR(1) coefficient A_z can be directly estimated using the estimated QR-score process $\hat{Z}_t = X_t\hat{m}_t$ with $\hat{m}_t := \tau - 1(\hat{e}_t \leq 0)$. However, as noted by Galvao and Yoon (2024), this procedure can induce substantial finite-sample biases in the autoregressive coefficients in $\hat{A}_z$, which is driven by the estimated non-smooth component $\hat{m}_t := (\tau - 1(\hat{e}_t \leq 0))$.

To avoid finite sample problems in $\hat{m}_t$, we follow the alternative formulation of $\hat{A}_z$ of Galvao and Yoon (2024) and extend their approach for the single-dimensional case to the general multivariate case, with $A_z \in \mathbb{R}^{d\times d}$. The alternative procedure assumes that the components X_t and e_t follow VAR(1) and Gaussian-AR(1) processes, respectively. The estimates for the VAR coefficient, $\hat{A}_x$, and the AR(1) coefficient for $\hat{m}_t$, $\hat{\phi}(\tau)$, are estimated separately. The former can be straightforwardly formulated by $\hat{A}_x = (\sum_{t=1}^T X_t X_{t-1}')(\sum_{t=1}^T X_{t-1}X_{t-1}')^{-1}$. For the latter component, recall that $(e_t, e_{t-1})'$ is assumed to follow a bivariate normal distribution. The implied AR(1) coefficient for $\hat{m}_t$ can be formulated as below: (i) Compute a normalized $\breve{e}_t = \hat{e}_t/s_e$, where $\hat{e}_t = y_t - X_t'\hat{\beta}(\tau)$, and s_e is the sample standard deviation for $\{\hat{e}_t\}_{t=1}^T$. (ii) Compute $\hat{\rho}_e = \sum_{t=2}^T(\breve{e}_t - \bar{e}_+)(\breve{e}_{t-1} - \bar{e}_-)/\sum_{t=2}^T(\breve{e}_{t-1} - \bar{e}_-)^2$, where $\bar{e}_+ = (T-1)^{-1}\sum_{t=2}^T \breve{e}_t$ and

$\bar{e}_{-} = (T-1)^{-1}\sum_{t=1}^{T-1}\breve{e}_t$. (iii) Compute $\hat{q}(\tau)$, the τ-quantile of $\{\breve{e}_t\}_{t=1}^{T}$ (iv) Obtain the alternative AR(1) coefficient estimator:

$$\hat{\phi}(\tau) = \frac{\Phi_2\left[(0,0)'; \mu, V\right] - \tau^2}{\tau(1-\tau)} \text{ with } \mu = \begin{pmatrix} -\hat{q}(\tau) \\ -\hat{q}(\tau) \end{pmatrix} \text{ and } V = \begin{pmatrix} 1 & \hat{\rho}_e \\ \hat{\rho}_e & 1 \end{pmatrix},$$

where $\Phi_2(\cdot; \mu, V) : \mathbb{R}^2 \to \mathbb{R}$ denotes the CDF of bivariate normal random variable $N(\mu, V)$.

With $\hat{A}_x$ and $\hat{\phi}(\tau)$, we can obtain the alternative estimates for the VAR(1) coefficient $\hat{A}_z = \hat{A}_x\hat{\phi}(\tau)$. We can then derive the implied parametric plugged-in estimates $\hat{\Omega}_{\text{var}}(\tau)$ and $\hat{B}_{\text{var}}(\tau)$ by replacing A_z and Υ in (53) and (54) with $\hat{A}_z = \hat{A}_x\hat{\phi}(\tau)$, and $\hat{\Upsilon} = T^{-1}\sum_{t=1}^{T}\hat{w}_{zt}\hat{w}_{zt}'$ with $\hat{w}_{zt} = \hat{Z}_t - \hat{A}_z\hat{Z}_{t-1}$, respectively.

8.2 Tables in Section 6

Table 1: Empirical rejection ratios for $\tau = 0.50$ with Normal QR errors and $\alpha = 0.05$

τ= 0.50 and $F_e \sim N(0,1)$ for innovation in (45)													
		$p=1$				$p=2$				$p=3$			
T	ρ	HAR-M	QR-HAC	QR-HAR	SETBB	HAR-M	QR-HAC	QR-HAR	SETBB	HAR-M	QR-HAC	QR-HAR	SETBB
200	0.250	0.072	0.081	0.070	0.020	0.074	0.097	0.081	0.014	0.086	0.114	0.089	0.010
200	0.500	0.141	0.095	0.073	0.040	0.183	0.123	0.091	0.030	0.206	0.152	0.109	0.025
200	0.750	0.347	0.119	0.089	0.091	0.464	0.157	0.116	0.074	0.537	0.186	0.138	0.075
200	0.900	0.563	0.134	0.092	0.179	0.743	0.176	0.110	0.179	0.686	0.208	0.127	0.198
400	0.250	0.069	0.071	0.063	0.020	0.076	0.076	0.066	0.013	0.079	0.089	0.073	0.012
400	0.500	0.139	0.078	0.064	0.036	0.185	0.090	0.069	0.025	0.215	0.101	0.085	0.025
400	0.750	0.370	0.091	0.070	0.081	0.517	0.106	0.082	0.064	0.613	0.132	0.095	0.070
400	0.900	0.649	0.100	0.075	0.155	0.809	0.119	0.084	0.164	0.844	0.143	0.094	0.189
800	0.250	0.074	0.063	0.058	0.026	0.086	0.072	0.063	0.019	0.090	0.075	0.063	0.016
800	0.500	0.147	0.079	0.067	0.043	0.203	0.083	0.069	0.030	0.232	0.086	0.077	0.029
800	0.750	0.384	0.084	0.068	0.079	0.539	0.095	0.074	0.067	0.649	0.104	0.078	0.066
800	0.900	0.679	0.087	0.069	0.138	0.871	0.102	0.080	0.150	0.933	0.110	0.091	0.167

Note: HAR-M reports empirical null rejection probabilities for conditional mean HAR-Wald inference using OLS estimates with F-criticalvalues. It employs the smoothing parameter K^*, which minimizes the testing-oriented selection rule according to the formula provided in Ye and Sun (2018). QR-HAC implements the HAC-based QR-Wald inference with kernel LRV estimates, employing the Bartlett kernel and chi-square critical values derived from conventional increasing-smoothing (or small-b) asymptotics. It employs the smoothing parameter b^*, which minimizes the asymptotic mean squared error (AMSE) according to the formula provided in Galvao and Yoon (2024). QR-HAR is based on the OS-LRV estimate, using the F-criticalvalues. The smoothing parameter $K^*$for W_{HAR} is determined using a testing-oriented rule with the data-dependent formulation specified in Section 5. SETBB refers to bootstrap-based inference (smooth extended tapered block bootstrap) for QR, proposed by Gregory et al. (2018). The nominal level of test α is set at 5%, with 5,000 replications.

Table 2: Empirical rejection ratios for $\tau = 0.50$ with Chi-square QR errors and $\alpha = 0.05$

τ= 0.50 and $F_e \sim (\chi_2(1) - 1)/\sqrt{2}$ for innovation in (45)													
		$p = 1$				$p = 2$				$p = 3$			
n	ρ	HAR-M	QR-HAC	QR-HAR	SETBB	HAR-M	QR-HAC	QR-HAR	SETBB	HAR-M	QR-HAC	QR-HAR	SETBB
200	0.250	0.064	0.085	0.074	0.033	0.085	0.103	0.085	0.020	0.103	0.119	0.098	0.016
200	0.500	0.121	0.097	0.075	0.047	0.167	0.127	0.094	0.031	0.200	0.154	0.114	0.027
200	0.750	0.320	0.123	0.095	0.088	0.428	0.163	0.117	0.065	0.482	0.197	0.142	0.068
200	0.900	0.549	0.132	0.087	0.165	0.727	0.180	0.110	0.169	0.660	0.224	0.127	0.189
400	0.250	0.066	0.079	0.069	0.057	0.080	0.090	0.076	0.040	0.101	0.101	0.084	0.033
400	0.500	0.135	0.090	0.076	0.061	0.192	0.102	0.081	0.045	0.236	0.127	0.106	0.040
400	0.750	0.363	0.100	0.076	0.093	0.504	0.128	0.099	0.073	0.601	0.152	0.110	0.071
400	0.900	0.627	0.112	0.084	0.153	0.795	0.136	0.092	0.151	0.787	0.163	0.111	0.172
800	0.250	0.074	0.068	0.060	0.060	0.087	0.073	0.061	0.048	0.109	0.082	0.070	0.048
800	0.500	0.149	0.072	0.063	0.060	0.199	0.085	0.070	0.051	0.241	0.099	0.092	0.048
800	0.750	0.388	0.088	0.071	0.080	0.542	0.106	0.091	0.075	0.652	0.121	0.088	0.074
800	0.900	0.679	0.093	0.073	0.135	0.855	0.110	0.087	0.139	0.905	0.125	0.102	0.154

See footnote in Table 1.

Table 3: Empirical rejection ratios for $\tau = 0.50$ with$T(3)$ QR errors and $\alpha = 0.05$

τ= 0.50 and $F_e \sim T_3/\sqrt{3}$ for innovation in (45)													
		$p = 1$				$p = 2$				$p = 3$			
T	ρ	HAR-M	QR-HAC	QR-HAR	SETBB	HAR-M	QR-HAC	QR-HAR	SETBB	HAR-M	QR-HAC	QR-HAR	SETBB
200	0.250	0.057	0.082	0.071	0.012	0.076	0.097	0.075	0.006	0.106	0.111	0.087	0.005
200	0.500	0.109	0.098	0.079	0.030	0.146	0.121	0.088	0.018	0.186	0.146	0.108	0.012
200	0.750	0.295	0.125	0.092	0.073	0.384	0.158	0.118	0.051	0.437	0.195	0.140	0.044
200	0.900	0.534	0.139	0.084	0.156	0.707	0.178	0.105	0.135	0.634	0.213	0.122	0.151
400	0.250	0.059	0.071	0.062	0.017	0.082	0.079	0.070	0.010	0.110	0.088	0.073	0.007
400	0.500	0.127	0.079	0.066	0.034	0.166	0.097	0.073	0.023	0.221	0.107	0.085	0.019
400	0.750	0.332	0.095	0.072	0.064	0.463	0.120	0.090	0.055	0.554	0.141	0.103	0.052
400	0.900	0.588	0.104	0.073	0.129	0.762	0.131	0.089	0.123	0.744	0.154	0.100	0.125
800	0.250	0.067	0.069	0.061	0.022	0.085	0.076	0.066	0.014	0.106	0.076	0.065	0.011
800	0.500	0.135	0.080	0.068	0.038	0.188	0.086	0.073	0.026	0.230	0.098	0.087	0.024
800	0.750	0.349	0.081	0.062	0.062	0.506	0.105	0.091	0.058	0.606	0.108	0.087	0.052
800	0.900	0.635	0.086	0.068	0.117	0.804	0.104	0.080	0.118	0.844	0.112	0.096	0.123

See footnote in Table 1.

Table 4: Empirical rejection ratios for $\tau \in \{0.75, 0.90\}$ with $p = 1$ and $\alpha = 0.05$

DGP in (49) with $p = 1$

		τ= 0.75			τ= 0.90		
T	ρ	QR-HAC	QR-HAR	SETBB	QR-HAC	QR-HAR	SETBB
200	0.250	0.086	0.076	0.027	0.109	0.102	0.025
200	0.500	0.113	0.098	0.051	0.129	0.116	0.049
200	0.750	0.141	0.108	0.091	0.178	0.152	0.087
200	0.900	0.192	0.136	0.165	0.246	0.194	0.146
400	0.250	0.081	0.074	0.022	0.094	0.087	0.019
400	0.500	0.087	0.077	0.039	0.115	0.106	0.031
400	0.750	0.114	0.094	0.078	0.139	0.120	0.069
400	0.900	0.152	0.104	0.131	0.185	0.156	0.128
800	0.250	0.078	0.069	0.023	0.079	0.076	0.015
800	0.500	0.084	0.086	0.036	0.092	0.087	0.026
800	0.750	0.092	0.080	0.060	0.105	0.094	0.050
800	0.900	0.114	0.089	0.101	0.139	0.117	0.101

See footnote in Table 1.

Table 5: Empirical rejection ratios for $\tau \in \{0.75, 0.90\}$ with $p = 2$ and $\alpha = 0.05$

DGP in (49) with $p = 2$

		τ= 0.75			τ= 0.90		
T	ρ	QR-HAC	QR-HAR	SETBB	QR-HAC	QR-HAR	SETBB
200	0.250	0.114	0.096	0.018	0.160	0.147	0.014
200	0.500	0.148	0.121	0.039	0.198	0.179	0.035
200	0.750	0.198	0.154	0.084	0.270	0.236	0.069
200	0.900	0.257	0.179	0.145	0.351	0.286	0.140
400	0.250	0.102	0.089	0.016	0.139	0.129	0.016
400	0.500	0.121	0.107	0.032	0.168	0.155	0.028
400	0.750	0.153	0.130	0.069	0.208	0.189	0.060
400	0.900	0.191	0.147	0.119	0.271	0.235	0.113
800	0.250	0.089	0.078	0.017	0.106	0.102	0.009
800	0.500	0.105	0.105	0.032	0.127	0.122	0.021
800	0.750	0.120	0.107	0.056	0.148	0.140	0.041
800	0.900	0.143	0.117	0.102	0.197	0.178	0.093

See footnote in Table 1.

8.3 Technical lemmas

The following lemma is a version of Lemma C.2 in Galvao and Kato (2016) which states a property of strong mixing process $\{V_t\}$ which provides upper bound for covariance between V_t and V_{t+s} for any fixed integer s. As noted in Galvao and Yoon (2024), the original result is attributed to Yoshihara (1976), who assumed β-mixing, but this result can be applied to the case of α-mixing in Lemma 1 below.

Lemma 1. *Let $\{V_t\}$ be with α-mixing coefficient $\alpha[s]$ such that $\mathbb{E}[V_t] = \mathbb{E}[V_{t+s}] = 0$, and for some positive constants δ and M such that*

$$\mathbb{E}[|V_t|^{1+\delta}]\mathbb{E}[|V_{t+s}|^{1+\delta}] \leq M \text{ and } \mathbb{E}[|V_t V_{t+s}|^{1+\delta}] \leq M.$$

Then, we have that

$$|cov(V_t, V_{t+s})| \leq 4M^{1/(1+\delta)}\alpha[s]^{\delta/(1+\delta)}.$$

The next lemma provides a slight modification of the exponential-type inequality result from equation (2.1) in Theorem 1 of Merlevéde, Peligrad, Rio (2009). The inequality in Lemma 2 extends Bernstein's inequality for i.i.d. data to time series which is characterized by α-mixing dependence.

Lemma 2. *Let $\{U_t\}_{t=1}^T$ be a sequence of zero-mean stationary random variables that satisfy*

$$\alpha[k] \leq \exp(-a_0 k)$$

for a certain constant $a_0 > 0$. Also, assume that there exists a positive constant B that satisfies $\max_{1\leq t\leq T} ||U_t|| \leq B$. Then, there exists a positive constant c_1 depending only on a_0 such that the following inequality holds for all $x \geq 0$:

$$P\left(\left|\sum_{t=1}^T U_t\right| \geq x\right) \leq \exp\left(-\frac{c_1 x^2}{\nu^2 + B^2 + xB(\log n)^2}\right),$$

for sufficiently large $T \geq 2$, where $\nu^2 = var(\sum_{t=1}^T U_t)$.

Our final technical result in this subsection shows that the CLT assumption in Assumption 5 can be validated under certain primitive conditions, similar to those in Jenish and Prucha (2009). For simplicity of exposition, we focus on the marginal weak convergence in the CLT of $\{\Phi_k(t/T) Z_t\}_{t=1}^T$ for a scalar time series Z_t, as the joint weak convergence over $k \in \{1, \ldots, J\}$ and the multivariate Z_t case can be established using the Cramér-Wold device. For a generic random variable V, we denote the upper quantile function $L_V(\cdot) : (0, 1) \to [0, \infty)$ as

$$L_V(u) = \inf\{x \in \mathbb{R} : P(V > x) \leq u\}.$$

Additionally, we define the inverse function of the α-mixing coefficient $\alpha_{inv}[u] : (0, 1) \to \mathbb{N} \cup \{0\}$, as:

$$\alpha_{inv}[u] = \max\{s \geq 0 : \alpha[s] > u\}.$$

Lemma 3. *Suppose that:*

i) There exists an array of positive real constants $\{c_{t,T}\}_{t=1}^{T}$ *such that*

$$\lim_{M\to\infty}\sup_{T\in\mathbb{N}}\sup_{1\le t\le T}\ \mathbb{E}\left[\left|\Phi_k\left(\frac{t}{T}\right)\left(\frac{Z_t}{c_{t,T}}\right)\right|^2 1\left(\left|\Phi_k\left(\frac{t}{T}\right)\left(\frac{Z_t}{c_{t,T}}\right)\right|>M\right)\right]=0;$$

$$\lim\inf_{T\to\infty}\left\{\frac{1}{T}\left(\frac{\Omega_{T,k}}{\max_{1\le t\le T}c_{t,T}}\right)\right\}>0,$$

where $\Omega_{T,k}(\tau):=var(T^{-1/2}\sum_{t=1}^{T}\Phi_k(t/T)Z_t)$.

ii) The sequence $\{\Phi_k(t/T)\,Z_t\}_{t=1}^{T}$ *is* α*-mixing and satisfies that*

$$\lim_{M\to\infty}\lim_{T\to\infty}\sup_{1\le t\le T}\int_0^1\alpha_{inv}^2[u]\left(L_{|\Phi_k(t/T)Z_t|1(\Phi_k(t/T)Z_t|>C)}\right)^2du=0;$$

$$\sum_{j=1}^{\infty}j\alpha[j]<0 \text{ and } \alpha(j)=O(j^{-2-\epsilon}) \text{ for some } \epsilon>0.$$

If $\Phi_k(\cdot)$ *is continuously differentiable such that* $\int_0^1\Phi_k^2(r)dr=1$*, and if* $\sum_{j=1}^{T}j|\mathbb{E}[Z_tZ_{t-j}]|<\infty$*, then*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)Z_t\xrightarrow{d}N\left(0,\Omega(\tau)\right),$$

as $T\to\infty$.

8.4 Technical results for the proofs of main results

We begin by presenting some technical results to establish (18) for the empirical process $\mathbb{G}_{k,T}(b)$. Define the following versions of the empirical processes:

$$\check{\mathbb{G}}_{k,T}(b):=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left(Z_t(b)-\mathbb{E}[Z_t(b)|X_t]\right);$$

$$\tilde{\mathbb{G}}_{k,T}(b):=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\left(\mathbb{E}[Z_t(b)|X_T]-\mathbb{E}[Z_t(b)]\right).$$

Lemma 4 below shows that

$$\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}||\check{\mathbb{G}}_{k,T}(b)-\check{\mathbb{G}}_{k,T}(\beta_0(\tau))||=o_p(1) \text{ and } \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}||\tilde{\mathbb{G}}_{k,T}(b)-\tilde{\mathbb{G}}_{k,T}(\beta_0(\tau))||=o_p(1),$$

from which the result in (18) follows by the triangle inequality.

Lemma 4. *Suppose Assumptions 1–4 hold, and fix any positive constant* $\mathcal{C}_0$ *for* $\epsilon_T=\mathcal{C}_0T^{-1/2}$ *in* $\mathcal{N}_{\epsilon_T}(\beta_0(\tau))$*. Then, as* n *grows to infinity, we have that*

$$\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}||\check{\mathbb{G}}_{k,T}(b)-\check{\mathbb{G}}_{k,T}(\beta_0(\tau))||=O_p\left(\frac{\log n}{n^{1/4}}\right); \tag{55}$$

$$\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}||\tilde{\mathbb{G}}_{k,T}(b)-\tilde{\mathbb{G}}_{k,T}(\beta_0(\tau))||=O_p\left(\frac{(\log n)^3}{n^{3/10}}\right), \tag{56}$$

hold uniformly over $k\in\{0\}\cup\mathbb{N}$.

Proof of Lemma 4. Without loss of generality, we let X_i be single-dimensional, i.e., $d = 1$, whenever it is convenient. Also, we assume that $\beta_0(\tau) = 0$ without loss of generality. We first prove the result in (55). Define

$$\varphi_t(b) := \left\{1(e_t \leq 0) - 1(e_t \leq X_t'b)\right\} - \left\{\mathbb{E}[1(e_t \leq 0)|X_t] - \mathbb{E}[1(e_t \leq X_t'b)|X_t]\right\};$$
$$\psi_t(b) := \varphi_t(b)X_t.$$

Then, it is not difficult to check that $\mathbb{E}[\Phi_k(t/T)\psi_t(b)] = 0$ for any $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$ and $k \in \mathbb{N} \cup \{0\}$, using the law of iterated expectations. Also, by construction, we have that

$$\begin{aligned}\breve{\mathbb{G}}_{k,T}(b) - \breve{\mathbb{G}}_{k,T}(\beta_0(\tau)) = \breve{\mathbb{G}}_{k,T}(b) - \breve{\mathbb{G}}_{k,T}(0) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\varphi_t(b)X_t \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b).\end{aligned}$$

With $\gamma_T = T^{-1/4}\log(T)$, we want to show that there exists a sufficiently large constant $M > 0$ that does not depend on $k \in \mathbb{N} \cup \{0\}$ and satisfy that

$$\lim_{T\to\infty} P\left(\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b)\right| > M\gamma_T\right) = 0. \tag{57}$$

To establish the result in (57), we follow the following steps 1–3. Step 1: Fix $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$, and show that

$$var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b)\right) = O\left(\frac{1}{\sqrt{T}}\right), \tag{58}$$

where the upper bound for the term $O(T^{-1/2})$ does not depend on k and b. From the stationarity of $\{\psi_t(b)\}$, we have that

$$\begin{aligned}var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b)\right) &= var\left(\Phi_k\left(\frac{t}{T}\right)\psi_1(b)\right) + 2\sum_{s=2}^{T}\left(1-\frac{t}{T}\right)Cov(\Phi_k\left(\frac{t}{T}\right)\psi_s(b), \Phi_k\left(\frac{1}{T}\right)\psi_1(b)) \\ &\leq \left(\sup_{1\leq s\leq T, k\in\mathbb{N}}\left|\Phi_k\left(\frac{s}{T}\right)\right|\right)^2 \times \left(var(\psi_1(b)) + 2\sum_{s=2}^{T}|Cov(\psi_s(b), \psi_1(b))|\right).\end{aligned} \tag{59}$$

For the variance term in (59), note that $var(\psi_1(b)) = \mathbb{E}[\psi_1^2(b)] = \mathbb{E}[X_1^2\mathbb{E}[\varphi_1^2(b)|X_1]]$ holds. Also, conditioning on X_1, $\varphi_t(b)$ is a centered Bernoulli random variable, and we denote its conditional success probability $p_b(X_1)$ for any $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. We then use the boundedness assumption $|f(e|x)| \leq C$ in Assumption 3 and obtain that

$$p_b(X_1) \leq |F(X_1'b|X_1) - F(0|X_1)| = \left|\int_{-\infty}^{X_1'b} f(e|X_1)de - \int_{-\infty}^{0} f(e|X_1)de\right| \leq (C\epsilon_T)\cdot|X_1| \tag{60}$$

for some constant $C > 0$. Using this result, we can bound the conditional variance of $\varphi_1^2(b)$, which is equal to $(1 - p_b(X_1))p_b(X_1)$, as below:

$$var(\varphi_1^2(b)|X_1] = \mathbb{E}[\varphi_1^2(b)|X_1] = (1 - p_b(X_1))p_b(X_1) \leq p_b(X_1) \leq (C\epsilon_T) \cdot |X_1|.$$

The results, together with Assumption 4-iv), lead us to conclude that

$$\mathbb{E}[\psi_1^2(b)] = \mathbb{E}[X_1^2 \mathbb{E}[\varphi_1^2(b)|X_1]] \leq (C\epsilon_T) \cdot \mathbb{E}[|X_i|^3] \leq C' \cdot \epsilon_T$$

holds for some constant $C' > 0$, and any $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. Next, we focus on covariance terms in (59). Consider $\mathbb{E}\left[\psi_{1+s}^2(b)\psi_1^2(b)\right] = \mathbb{E}\left[\varphi_{1+s}^2(b)\varphi_1^2(b)X_{1+s}^2 X_1^2\right]$. Conditioning on (X_{1+s}, X_1), $\varphi_1(b)$ and $\varphi_{1+s}(b)$ are centered Bernoulli random variables with success probabilities $p_b(X_1)$ and $p_b(X_{1+s})$, respectively. By (60), these probabilities are bounded by $\epsilon_T \cdot |X_t|$ and $\epsilon_T \cdot |X_{t+s}|$, respectively. Additionally, $\varphi_1(b)\varphi_{1+s}(b)$ is also a Bernoulli random variable with conditional success probability $p_b(X_t)p_b(X_{t+s})$. We then have that

$$\begin{aligned}
\mathbb{E}\left[\varphi_{1+s}^2(b)\varphi_1^2(b)|(X_{1+s}, X_1)\right] &= var(\varphi_1(b)\varphi_{1+s}(b)|(X_{1+s}, X_1)] + \mathbb{E}\left[\varphi_{1+s}(b)|(X_{1+s}, X_1)\right] \cdot \mathbb{E}\left[\varphi_1(b)|(X_{1+s}, X_1)\right] \\
&= p_b(X_1)p_b(X_{1+s}) \cdot (1 - p_b(X_1)p_b(X_{1+s})) + p_b(X_1)p_b(X_{1+s}) \\
&\leq 2p_b(X_1)p_b(X_{1+s}) \leq (2C^2\epsilon_T^2) \cdot |X_1 X_{1+s}|,
\end{aligned}$$

and this leads us to obtain that

$$\mathbb{E}\left[\psi_{1+s}^2(b)\psi_1^2(b)\right] = \mathbb{E}\left[\varphi_{1+s}^2(b)\varphi_1^2(b)X_{1+s}^2 X_1^2\right] \leq (2C^2\epsilon_T^2) \cdot \mathbb{E}[|X_{1+s}^3 X_1^3|]] \leq C''\epsilon_T^2$$

holds for some positive constant C'' that does not depend on s. Summing up the results so far, we verified that

$$\mathbb{E}[\psi_1^2(b)] = \mathbb{E}[\psi_{1+s}^2(b)] \leq C' \cdot \epsilon_T \text{ and } \mathbb{E}\left[\psi_{1+s}^2(b)\psi_1^2(b)\right] \leq C'' \cdot \epsilon_T^2$$

hold for some constants C' and C''. This finding allow us to apply Lemma 1 by setting $\delta = 1$ and $V_t = \psi_t(b)$, and we obtain that

$$|Cov(\psi_{1+s}(b), \psi_1(b))| \leq C'''\epsilon_T \alpha[s]^{1/2}$$

holds for some universal constant $C''' > 0$. Thus, we conclude that

$$\sum_{s=2}^{T} |Cov(\psi_s(b), \psi_1(b))| \leq \sum_{s=1}^{T} |Cov(\psi_s(b), \psi_1(b))| \leq C''' \cdot \epsilon_T \cdot \left(\sum_{s=1}^{\infty} \alpha[s]^{1/2}\right),$$

where the second-to-last inequality follows from Assumption 2, $\sum_{s=1}^{\infty} \alpha[s]^{1/2} < \infty$. This implies that the following inequality

$$var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(b)\right) \leq var(\psi_1(b)) + 2\sum_{s=2}^{T}|Cov(\psi_s(b), \psi_1(b))| \leq \frac{\tilde{C}}{\sqrt{T}} \tag{61}$$

holds for sufficiently large T with some positive $\tilde{C} > 0$. This result, together with uniform boundedness of $\Phi_k(\cdot)$, implies that

$$var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b)\right) \leq \left(\sup_{1\leq s\leq T, k\in\mathbb{N}}\left|\Phi_k\left(\frac{s}{T}\right)\right|\right)^2 \times \left(var(\psi_1(b)) + 2\sum_{s=2}^{T}|Cov(\psi_s(b),\psi_1(b))|\right) \tag{62}$$

$$= O\left(\frac{1}{\sqrt{T}}\right),$$

which is the desired result. Step 2: Fixed $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. Using the result in Step 1, we show that there exist some positive constants c_1 and $\tilde{C}$ such that

$$P\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b)\right| > M\gamma_T\right) = \exp\left(-\frac{c_1}{\tilde{C}}M^2\log(T)^2\right) \tag{63}$$

holds for any finite $M > 0$ that does not depend on $k \in \mathbb{N}\cup\{0\}$. We apply Lemma 2 to $U_t = \Phi_k(t/T)\psi_t(b)$, $B = \Delta\cdot O(T^{1/5})$, $\gamma_T = T^{-1/4}\log T$, and $x = M\sqrt{T}\gamma_T$, and obtain that, for sufficiently large T,

$$\begin{aligned}
P\left(\left|\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b)\right| > M\sqrt{T}\gamma_T\right) &\leq \exp\left(-\frac{c_1\left(M\sqrt{T}\gamma_T\right)^2}{var\left(\sum_{t=1}^{T}\psi_t(b)\right) + \left(M\sqrt{T}\gamma_T\right)B\log(T)^2 + B^2}\right)\\
&\leq \exp\left(-\frac{c_1M^2\left(\log T\right)^2T^{1/2}}{\tilde{C}T^{1/2} + MT^{1/4}(\log T)^3B + B^2}\right)\\
&= \exp\left(-\frac{c_1M^2\left(\log T\right)^2}{\tilde{C} + \Delta Mn^{-1/4}O(n^{1/5})\log(T)^3 + \Delta^2T^{2/5}T^{-1/2}}\right)\\
&= \exp\left(-\frac{c_1M^2\left(\log T\right)^2}{\tilde{C} + O(T^{-1/20})\log(T)^3 + O(T^{-1/10})}\right)
\end{aligned}$$

holds for any fixed $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. The result implies that

$$P\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b)\right| > M\gamma_T\right) \leq \exp\left(-\frac{c_1}{\tilde{C}}M^2\log(T)^2\right)$$

holds for constants $c_1, \tilde{C} > 0$ and sufficiently large number of T. Step 3: We now extend the pointwise result of (63) in Step 2 to the uniform one with respect to all $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. This can be shown by extending a classical chaining technique, e.g., Bickel (1975) and Zhou and Portnoy (1996), and accounting for the temporal dependence in the process $\{\psi_t(b)\}_{t=1}^{T}$. Let $\ell_T = T^{-1/4}$. We first partition the local parameter space $\mathcal{N}_{\epsilon_T}(\beta_0(\tau))$ into $\cup_{j=1}^{N}(\mathcal{N}_{\epsilon_T}(\beta_0)\cap E_j)$, where E_j, for $j \in \{1,\ldots,N\}$, are closed cubes, whose vertices are on the set

$$\{(k_{j_1}\ell_T\epsilon_T,\ldots,k_{j_d}\ell_T\epsilon_T)\} \text{ for } k_{j_m} \in \{0,\pm1,\ldots,\pm\left\lfloor\ell_T^{-1}\right\rfloor+1\}$$

with $m \in \{1,\ldots,d\}$. If $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))\cap E_j$, we denote b_j as the lowest vertex of the cube E_j containing b. The side length of each cube E_j is $\ell_T\epsilon_T = \mathcal{C}_0T^{-3/4}$, and we set the total number of cubes equal to

$N = (2\lfloor \ell_T^{-1} \rfloor + 3)^d$. We have that

$$\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) \psi_t(b) \right| \leq \max_{1 \leq j \leq N} P_{k,T}(b_j) + \max_{1 \leq j \leq N} Q_{k,T}(b_j),$$

where

$$P_{k,T}(b_j) = \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) \psi_t(b_j) \right|;$$
$$Q_{k,T}(b_j) = \sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau)) \cap E_j} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) (\psi_t(b) - \psi_t(b_j)) \right|.$$

Without loss of generality, we assume that ℓ_T^{-1} is an integer so that $N = (2\ell_T^{-1} + 3)^d$. By Bonferroni inequality and the result in step 2, we have that

$$\begin{aligned}
P\left(\max_{1 \leq j \leq N} P_{k,T}(b_j) > M\gamma_T\right) &\leq N \cdot \max_{1 \leq j \leq N} P\left(P_{k,T}(b_j) > M\gamma_T\right) \\
&\leq (2\ell_T^{-1} + 3)^d \cdot \exp\left(-\frac{c_1}{\tilde{C}} M^2 (\log T)^2\right) \\
&\leq (2T^{1/4} + 3)^d \cdot \exp\left(-\frac{c_1}{\tilde{C}} M^2 (\log T)^2\right) \\
&\leq \max\{1, 2^{d-1}\} \cdot (2^d T^{d/4} + 3^d) \cdot \exp\left(-\frac{c_1}{\tilde{C}} M^2 (\log T)^2\right) \\
&\leq \tilde{C} \cdot \exp\left(\frac{d}{4} \log T\right) \cdot \exp\left(-\frac{c_1}{\tilde{C}} M^2 (\log T)^2\right) \\
&\leq \tilde{C} \cdot \exp\left(\frac{d}{4} \log T - \frac{c_1}{\tilde{C}} M (\log T)^2\right) = o\left(\frac{1}{T}\right).
\end{aligned}$$

holds for some constant $\tilde{C}$ and any finite $M > 0$. Next, we consider the term $Q_{k,T}(b)$. Given the positive sequence $a_{t,T} = |X_t| \ell_T \epsilon_T$, let us denote that

$$\begin{aligned}
\tilde{\varphi}_t(b, a_{t,T}) :=& \{1(e_t \leq X_t' b + a_{t,T}) - 1(e_t \leq X' b - a_{t,T})\} \\
& - \left\{\mathbb{E}[1(e_t \leq X_t' b + a_{t,T}) | X_t] - \mathbb{E}[1(e_t \leq X' b - a_{t,T}) | X_t]\right\}; \\
\tilde{\psi}_t(b, a_{t,T}) :=& |X_t| \tilde{\varphi}_t(b, a_{t,T}).
\end{aligned}$$

By Assumption 4-ii) and $|\tilde{\varphi}_t(b, a_{t,T})| \leq 1$ we have that a.s., $\max_{1 \leq t \leq T} |\tilde{\psi}_t(b, a_{t,T})| \leq \Delta T^{1/5}$. Given $j \in \{1, \ldots, N\}$, we utilize the monotonicity of the indicator function and obtain that

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) (\psi_t(b_j) - \psi_t(b)) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) X_t \left(1(e_t \leq X_t' b) - 1(e_t \leq X_t' b_j) - \left\{\mathbb{E}[1(e_t \leq X_t' b | X_t] - \mathbb{E}[1(e_t \leq X_t' b_j | X_t]\right\}\right) \\
&\leq \sup_{1 \leq s \leq T} \left|\Phi_k\left(\frac{s}{T}\right)\right| \times \left\{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} |X_t| \left\{1(e_t \leq X_t' b_j + a_{t,T}) - 1(e_t \leq X' b_j - a_{t,T})\right\}\right. \\
&\quad \left. + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} |X_t| \left\{\mathbb{E}[1(e_t \leq X_t' b_j + a_{t,T}) | X_t] - \mathbb{E}[1(e_t \leq X_t' b_j - a_{t,T} | X_t])\right\}\right\},
\end{aligned}$$

where the inequality holds for all $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau)) \cap E_j$. By triangle inequality, we can further bound the right side of the inequality by

$$\left(\sup_{1 \le s \le T, k \in \mathbb{N}} \left|\Phi_k\left(\frac{s}{T}\right)\right|\right) \cdot (I_{j,1} + I_{j,2}), \tag{64}$$

where

$$\begin{aligned}
I_{j,1} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} |\tilde{\psi}_t(b_j, a_{t,T});\\
I_{j,2} &= \frac{2}{\sqrt{T}} \sum_{t=1}^{T} |X_t| \left\{\mathbb{E}[1(e_t \le X_t' b_j + a_{t,T})|X_t] - \mathbb{E}[1(e_t \le X_t' b_j - a_{t,T}|X_t])\right\}.
\end{aligned}$$

We want to show that

$$\max_{1 \le j \le N} |I_{j,1}| = O_p\left(\frac{(\log T)^3}{T^{3/10}}\right) \text{ and } \max_{1 \le j \le N} |I_{j,2}| = \left(\frac{\log T}{T^{1/4}}\right). \tag{65}$$

For all $j \in \{1, \ldots, N\}$, we can check that $I_{j,1}$ has a zero mean using the law of iterated expectation. Additionally, it follows that

$$var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) \tilde{\psi}_t(b_j, a_{t,T})\right) \le \breve{C} \cdot \left(\frac{1}{T^{3/4}}\right) \tag{66}$$

holds for some $\breve{C}$ and any fixed $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. The comprehensive proof for (66) can be conducted similarly to that of (58) in Step 1, by showing that the variance of $I_{j,1}$ is uniformly bounded by $O(\ell_T \epsilon_T) = O(T^{-3/4})$. We omit the details. Now, set $\eta_T = T^{-3/10}(\log T)^3$, and we obtain that, for any $M > 0$,

$$P\left(\max_{1 \le j \le N} |I_{j,1}| \ge M\eta_T\right) \le \sum_{j=1}^{N} P\left(|I_{j,1}| > M\eta_T\right).$$

For the summand in the right hand side of the inequality, we apply Lemma 2 to $U_t = \Phi_k(t/T)\tilde{\psi}_t(b_j, a_{t,T})$, $B = \Delta T^{1/5}$, $\eta_T = T^{-3/10}(\log T)^3$, and $x = M\sqrt{T}\eta_T$, and obtain that, for sufficiently large T,

$$\begin{aligned}
P\left(|I_{j,1}| > M\eta_T\right) &= P\left(\left|\sum_{t=1}^{T} \tilde{\psi}_t(b_j, a_{t,T})\right| > M\sqrt{T}\eta_T\right)\\
&\le \exp\left(-\frac{c_1\left(M\sqrt{T}\eta_T\right)^2}{var(\sum_{t=1}^{T}\tilde{\psi}_t(b_j, a_{t,T})) + (M\sqrt{T}\eta_T)B(\log T)^2 + B^2}\right)\\
&= \exp\left(-\frac{c_1\left(M\sqrt{T}(\log T)^3 T^{-3/10}\right)^2}{var(\sum_{t=1}^{T}\tilde{\psi}_t(b_j, a_{t,T})) + (M\sqrt{T}(\log T)^3 T^{-3/10})B(\log T)^2 + B^2}\right)\\
&= \exp\left(-\frac{c_1 M^2 (\log T)^6 T^{2/5}}{\breve{C}T^{1/4} + \Delta M(\log T)^5 T^{2/5} + \Delta^2 T^{2/5}}\right)\\
&= \exp\left(-\frac{c_1 M^2(\log T)}{\breve{C}T^{-3/20}(\log T)^{-5} + \Delta M + \Delta^2(\log T)^{-5})}\right)\\
&\le \exp\left(-\frac{c_1 M}{\Delta}\log(T)\right).
\end{aligned}$$

Note that the upper bound does not depend on $j \in \{1, \dots, N\}$. Thus, for $M > 0$ sufficiently large, we have that

$$
\begin{aligned}
P\left(\max_{1\leq j\leq N} |I_{j,1}| \geq M\eta_T\right) &\leq \sum_{j=1}^{N} P\left(|I_{j,1}| > M\eta_T\right) \leq N \cdot \exp\left(-\frac{c_1 M}{\Delta}\log(T)\right) \\
&\leq (2T^{1/4}+3)^d \cdot \exp\left(-\frac{c_1 M}{\Delta}\log(T)\right) \\
&\leq \tilde{C}' \cdot \exp\left(\frac{d}{4}\log T\right)\exp\left(-\frac{c_1 M}{\Delta}\log(T)\right) \\
&\leq \tilde{C}' \cdot \exp\left(\left(\frac{d}{4} - \frac{c_1 M}{\Delta}\right)\log T\right) = o\left(\frac{1}{T}\right),
\end{aligned}
$$

for some $\tilde{C}' > 0$, where the last equation follows by choosing sufficiently large $M > (d/4)\cdot(\Delta/c_1)$. This shows the first result in (65). For the second result in (65), we have that

$$
\begin{aligned}
I_{j,2} &= \frac{2}{\sqrt{T}}\sum_{t=1}^{T} ||X_t|| \cdot \int_{-\infty}^{X_t' b_j + a_{t,T}} f(e|X_t)de - \int_{-\infty}^{X_t' b_j - a_{t,T}} f(e|X_t)de \\
&\leq \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left(||X_t|| \cdot \int_{X_t' b_j - |X_t|\ell_T\epsilon_T}^{X_t' b_j + |X_t|\ell_T\epsilon_T} f(e|X_t)de\right).
\end{aligned}
$$

Note that

$$
|X_t' b_j| \leq ||X||_\infty \cdot \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||b|| = ||X||_\infty \epsilon_T = \frac{\Delta C_0}{T^{3/10}},
$$

and this leads us to obtain that

$$
\begin{aligned}
\int_{X_t b_j - |X_t|\ell_T\epsilon_T}^{X_t b_j + |X_t|\ell_T\epsilon_T} f(e|X_t)de &\leq \sup_{|q|\leq ||X||_\infty\epsilon_T} \int_{q-||X_t||\ell_T\epsilon_T}^{q+||X_t||\ell_T\epsilon_T} f(e|X_t)de \\
&\leq ||X_t||\ell_T\epsilon_T \sup_{\substack{|q|\leq ||X_t||_\infty\epsilon_T \\ |h|\leq ||X_t||_\infty\ell_T\epsilon_T}} \frac{F(q+h|X_t) - F(q+h|X_t)}{|h|} \\
&\leq ||X_t||\ell_T\epsilon_T \sup_{\substack{|q|\leq ||X_t||_\infty\epsilon_T \\ |h|\leq ||X_t||_\infty\ell_T\epsilon_T}} \frac{F(q+h|X_t) - F(q+h|X_t)}{|h|}. \qquad (67)
\end{aligned}
$$

Since $||X_t||_\infty \ell_T \epsilon_T = O(T^{-11/20})$ is $o(1)$, $|h|$ in (67) shrinks to zero, as T increases to infinity. This, together with Assumption 3-i), implies that

$$
|I_{j,2}| \leq (\Delta \mathcal{C}_0 \ell_T) \cdot \left(\frac{2}{T}\sum_{t=1}^{T} ||X_t||^2\right) \cdot O(1) = O_p\left(\frac{1}{T^{1/4}}\right),
$$

where the upper bound does not depend on the index of cube $j \in \{1, \dots, N\}$. In summary so far, we have shown that

$$
\begin{aligned}
\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\check{\mathbb{G}}_{k,T}(b) - \check{\mathbb{G}}_{k,T}(\beta_0(\tau))|| &= \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))} \left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right)\psi_t(b)\right| \\
&\leq \max_{1\leq j\leq N} P_{k,T}(b_j) + \max_{1\leq j\leq N} Q_{k,T}(b_j), \qquad (68)
\end{aligned}
$$

and the upper bound in (68) satisfies that

$$\max_{1\le j\le N} P_{k,T}(b) = \max_{1\le j\le N}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\psi_t(b_j)\right| = O_p\left(\frac{\log T}{T^{1/4}}\right), \tag{69}$$

and

$$\begin{aligned}\max_{1\le j\le N} Q_{k,T}(b_j) &= \max_{1\le j\le N}\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))\cap E_j}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)(\psi_t(b)-\psi_t(b_j))\right| \\ &\le \sup_{1\le t\le T, k\in\mathbb{N}\cup\{0\}}\left|\Phi_k\left(\frac{t}{T}\right)\right|\cdot\left(\max_{1\le j\le N}|I_{j,1}| + \max_{1\le j\le N}|I_{j,2}|\right) \qquad (70) \\ &= O_p\left(\frac{(\log T)^3}{T^{3/10}}\right) + O_p\left(\frac{1}{T^{1/4}}\right) = O_p\left(\frac{\log T}{T^{1/4}}\right), \qquad (71)\end{aligned}$$

where the upper bounds for (69) and (71) do not depend on $k\in\mathbb{N}\cup\{0\}$. These results lead us to conclude that

$$\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\check{\mathbb{G}}_{k,T}(b) - \check{\mathbb{G}}_{k,T}(\beta_0(\tau))|| = O_p\left(\frac{\log T}{T^{1/4}}\right),$$

which is the desired result. To prove the result in (56), we define a mean zero process

$$\begin{aligned}\pi_t(b) &= (X_t\mathbb{E}[1(e_t\le 0) - 1(e_t\le X_t'b)|X_t]) - \mathbb{E}\left[X_t\left\{1(e_t\le 0) - 1(e_t\le X_t'b)\right\}\right] \\ &= X_t\tilde{p}_b(X_t) - \mathbb{E}\left[X_t\tilde{p}_b(X_t)\right]\end{aligned}$$

for any given $b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))$, where $\tilde{p}_b(X_t) = \mathrm{sgn}(X_tb)\cdot(F(X_1'b|X_t) - F(0|X_t))$. By construction, we have that

$$\begin{aligned}\tilde{\mathbb{G}}_{k,T}(b) - \tilde{\mathbb{G}}_{k,T}(\beta_0(\tau)) = \tilde{\mathbb{G}}_{k,T}(b) - \tilde{\mathbb{G}}_{k,T}(0) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\pi_t(b)X_t \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\xi_t(b).\end{aligned}$$

Then, the proof of (56) can be carried out in a similar manner to the proof of (55) by properly modifying its Steps 1–3. To be more specific, we follow Step 1'–3' as below: Step 1': For any fixed $b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))$, we have that

$$var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\xi_t(b)\right) = O\left(\frac{1}{n}\right), \tag{72}$$

where the upper bound for the term $O(T^{-1})$ does not depend on k and b. To prove this result, note that

$$\begin{aligned}var(\xi_t(b)) &= \mathbb{E}[X_t^2(F(X_tb|X_i) - F(0|X_t))^2] \\ &= \mathbb{E}[X_t^2(F(X_tb|X_t) - F(0|X_t))^2] \\ &\le \mathbb{E}[X_t^4]\cdot\epsilon_T^2 = O\left(\frac{1}{T}\right), \qquad (73)\end{aligned}$$

where the inequalities in (73) follows by Assumptions 3 and 4. Also, it can be shown that the (absolute) sum of covariance between $\xi_1(b)$ and $\xi_{1+s}(b)$ over $s \in \{2, \ldots, T\}$, is $O(T^{-1})$. Specifically, we can show that

$$\mathbb{E}\left[\xi_1^2(b)\right] = \mathbb{E}\left[\xi_{1+s}^2(b)\right] \le C'\epsilon_T^2;$$
$$\mathbb{E}[\xi_1^2(b)\xi_{1+s}^2(b)] \le C''\epsilon_T^4$$

for constants C' and C''. This allows us to apply the covariance inequality for mixing processes in Lemma 1 and obtain that

$$\sum_{s=2}^{T} |Cov(\psi_{1+s}(b), \psi_1(b))| \le O\left(\epsilon_T^2 \sum_{s=1}^{\infty} \alpha[s]^{1/2}\right) = O\left(\frac{1}{T}\right),$$

which can be completed in a similar way as shown in Step 1 of the proof of Lemma 4, we skip the details. Step 2': Choose $\gamma_T = (\log T)^3/T^{(3/10)}$. Then, using the result in Step 1', we show that there exists a positive constant M for sufficiently large T such that

$$P\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)\xi_t(b)\right| > M\gamma_T\right) = o\left(\frac{1}{T}\right). \tag{74}$$

holds for any fixed $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. Similar to what was shown in Step 2 of the proof of (63), the result in (74) can be verified by applying Lemma 2. We skip the details. Step 3': Extend the pointwise result of (63) in Step 2' to the uniform result with respect to all $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$, so that the result in (56) holds. This process can also be shown using the same technique shown in Step 3 of the proofs for (68)–(71), and the details are omitted. ■

To introduce the next lemma, we define

$$\check{\mathbb{M}}_T(b) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_T(t)(Z_t(b) - \mathbb{E}[Z_t(b)|X_t]);$$
$$\tilde{\mathbb{M}}_T(b) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_T(t)(\mathbb{E}[Z_t(b)|X_t] - \mathbb{E}[Z_t(b)]),$$

over $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$, where the weight function in these empirical processes is defined as

$$e_T(t) = T^{-1}\sum_{s=1}^{T} Q_h(t/T, s/T) - \int_0^1 Q_h\left(t/T, s\right) ds.$$

Lemma 5. *Suppose Assumptions 1–4 hold, and fix any positive constant $\mathcal{C}_0$ for $\epsilon_T = \mathcal{C}_0 T^{-1/2}$ in $\mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. Then, as T grows to infinity, we have that*

$$\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\check{\mathbb{M}}_T(b) - \check{\mathbb{M}}_T(\beta_0(\tau))|| = o_p\left(\frac{\log T}{T^{1/4}}\right); \tag{75}$$
$$\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\tilde{\mathbb{M}}_T(b) - \tilde{\mathbb{M}}_T(\beta_0(\tau))|| = o_p\left(\frac{(\log T)^3}{T^{3/10}}\right). \tag{76}$$

Proof of Lemma 5. From Assumption 1, where the functions used to construct $Q_h(\cdot,\cdot)$ are piecewise-monotonic, it follows that $\sup_{1\le s\le T}|e_T(s)| = O(1/T)$ holds for any fixed h. Using this property, proofs of (75) and (76) can be followed in exactly the same manner as the proofs of (55) and (56) in Lemma 4. The main difference involves replacing $\Phi_k(\cdot)$ in $\check{\mathbb{G}}_{k,T}(b)$ and $\tilde{\mathbb{G}}_{k,T}(b)$ with $e_T(\cdot)$ in $\check{\mathbb{M}}_T(b)$ and $\tilde{\mathbb{M}}_{k,T}(b)$. Specifically, we can replace $\sup_{1\le s\le T,k\in\mathbb{N}}|\Phi_k(s/T)| = O(1)$ with $\sup_{1\le s\le T}|e_T(s)| = O(1/T)$ in (59), (62), (64), and (70). This substitution allows us to obtain the desired convergence rates in (75) and (76). To save space, the details of these steps are omitted. ■

Lemma 6. *Let us denote $\tilde{\Omega}_h^*(\tau)$ as the infeasible version of $\hat{\Omega}_h^*(\tau)$, which is defined as:*

$$\tilde{\Omega}_h^*(\tau) = \frac{1}{n}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h^*\left(\frac{t}{T},\frac{s}{T}\right) Z_t Z_s'.$$

Then, as $T\to\infty$ and h is fixed, we have that

$$\tilde{\Omega}_h^*(\tau) \xrightarrow{d} \Omega^{1/2}\left(\sum_{k=1}^{\infty}\lambda_k \mathbb{Z}_k\mathbb{Z}_k'\right)\Omega^{1/2\prime},$$

where $\mathbb{Z}_k \overset{i.i.d}{\sim} N(0, I_d)$.

$$\begin{aligned} &P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right) Z_t \le x \text{ for } k = 0,1,...,J\right) \\ &= P\left(\Omega^{1/2}(\tau)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right) u_t \le x \text{ for } k=0,1,...,J\right) + o(1), \end{aligned}$$

for every fixed $J\in\mathbb{N}$, $x\in\mathbb{R}^d$ where $\Phi_0(\cdot) = 1$, $u_t \overset{i.i.d.}{\sim} N(0, I_d)$, and $\Omega^{1/2}(\tau)$ is a matrix square root of $\Omega(\tau)$ such that $\Omega^{1/2}(\tau)\Omega^{1/2}(\tau)' = \Omega(\tau)$.

Proof of Lemma 6. Recall that the spectral representation of $Q_h^*(\cdot,\cdot)$ in (15) enables us to re-express $\tilde{\Omega}_h^*(\tau)$ as:

$$\tilde{\Omega}_h^*(\tau) = \sum_{k=1}^{\infty}\lambda_k\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)Z_t\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Phi_k\left(\frac{s}{T}\right)Z_s\right)'.$$

For every $J\in\mathbb{N}$ fixed, define

$$\tilde{\Omega}_{h,J}^*(\tau) := \sum_{k=1}^{J}\lambda_k\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)Z_t\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Phi_k\left(\frac{s}{T}\right)Z_s\right)'.$$

Without loss of generality, we assume that $\tilde{\Omega}_h^*(\tau)$ and $\tilde{\Omega}_{h,J}^*(\tau)$ are scalar random variables. Then, we can apply the same truncation argument as in the proof of Lemma 1 in Sun (2014b). Setting $\omega_T = \tilde{\Omega}_h^*(\tau)$, $\xi_{T,J} = \tilde{\Omega}_{h,J}^*(\tau)$, and

$$\xi_{T,J}^* := \Omega^{1/2}(\tau)\sum_{k=1}^{J}\lambda_k\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_k\left(\frac{t}{T}\right)u_t\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Phi_k\left(\frac{s}{T}\right)u_s\right)\Omega^{1/2}(\tau)$$

for $u_t \overset{\text{i.i.d.}}{\sim} N(0, I_d)$, Assumptions 1 and 5 imply that conditions i)-iv) in Lemma A.3 of Sun (2014b) hold, including

$$\sup_T P\left(|\tilde{\Omega}_h^*(\tau) - \tilde{\Omega}_{h,J}^*(\tau)| > \delta\right) \xrightarrow{p} 0,$$

as $J \to \infty$, for every $\delta > 0$. Therefore, the result in Lemma A.3 of Sun (2014b) holds, which indicates

$$\begin{aligned}\tilde{\Omega}_h^*(\tau) &\xrightarrow{d} \Omega^{1/2}(\tau)\left(\int_0^1\int_0^1 Q_h^*(r,s)\,dW_d(r)dW_d(s)\right)\Omega^{1/2}(\tau)' \\ &\overset{d}{=} \Omega_{h,\infty}(\tau) := \Omega^{1/2}(\tau)\mathbb{S}_{h,\infty}\Omega^{1/2}(\tau)',\end{aligned}$$

as the desired result. ■

8.5 Proofs of main results

Proof of Theorem 1. Without loss of generality, we assume that the dimension of X_i is equal to one whenever it it is convenient. We first show the result in (13), whenever it is convenient. With simple algebra, we can write that

$$\begin{aligned}\hat{\Omega}_h^*(\tau) - \hat{\Omega}_h(\tau) &= \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\left\{Q_h^*\left(\frac{t}{T},\frac{s}{T}\right) - Q_{T,h}^*\left(\frac{t}{T},\frac{s}{T}\right)\right\}\hat{Z}_t\hat{Z}_s' \\ &= A_1 + A_2 - A_3,\end{aligned}$$

where

$$A_1 = \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T e_T(s)\hat{Z}_t\hat{Z}_s' \text{ and } A_2 = \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T e_T(t)\hat{Z}_t\hat{Z}_s',$$

with

$$e_T(t) = \frac{1}{T}\sum_{s=1}^T Q_h\left(\frac{t}{T},\frac{s}{T}\right) - \int_0^1 Q_h\left(\frac{t}{T},s\right)ds,$$

and

$$A_3 = \left[\frac{1}{T^2}\sum_{\tilde{t}=1}^T\sum_{\tilde{s}=1}^T Q_h\left(\frac{\tilde{t}}{T},\frac{\tilde{s}}{T}\right) - \int_0^1\int_0^1 Q_h(r,s)drds\right]\left[\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\hat{Z}_t\hat{Z}_s'\right].$$

Given that $T^{-1/2}\sum_{t=1}^T\hat{Z}_t = o_p(1)$ and $Q_h(\cdot,\cdot)$ is piecewise continuous and bounded over $[0,1]^2$, it is straightforward to verify that $A_3 = o_p(1)$. Next, we want to show that both $A_1 = o_p(1)$ and $A_2 = o_p(1)$. We first consider the term A_1, which can be bounded as below:

$$||A_1|| \leq \underbrace{\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T\hat{Z}_t\right\|}_{=o_p(1)\ \because(1)}\cdot\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^T e_T(s)\hat{Z}_s\right\|. \tag{77}$$

For the second term in the product on the right-hand side of (77), we have that

$$\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T} e_T(s)\hat{Z}_s\right\| \leq \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T} e_T(s)Z_s(b)\right\|$$

$$\leq \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T} e_T(s)(Z_s(b)-Z_s(\beta_0(\tau)))\right\| + \left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T} e_T(s)Z_s(\beta_0(\tau))\right\|, \quad (78)$$

where $Z_s(b) = X_s(\tau - 1(e_s \leq X_s(b-\beta_0(\tau))))$ for any $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. By construction, $Z_s(\beta_0(\tau))$ is equal to $Z_s = X_s(\tau - 1(e_s \leq 0))$. We use triangle inequality to construct the upper bound for the first part of (78):

$$\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T} e_T(s)(Z_s(b)-Z_s(\beta_0(\tau)))\right\|$$

$$\leq \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left\|\check{\mathbb{M}}_T(b)-\check{\mathbb{M}}_T(\beta_0(\tau))\right\| + \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left\|\tilde{\mathbb{M}}_T(b)-\tilde{\mathbb{M}}_T(\beta_0(\tau))\right\| \quad (79)$$

$$+ \left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T} e_T(s)\right\| \cdot \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left\|\mathbb{E}[Z_s(b)-Z_s(\beta_0)]\right\|,$$

where the terms in (79) are $o_p(1)$ from Lemma 5. To deal with the last term on the right-hand side of the inequality above, we use Assumption 3-i) and Taylor expansion conditioning on X_s, and obtain that

$$F(X_s'(b-\beta_0(\tau))|X_s) = F(0|X_s)+f(0|X_s)X_s(b-\beta_0(\tau))+f'(0|X_s)X_s^2(b-\beta_0(\tau))^2+f''(X_s\tilde{b}\,|X_s)X_s^3(b-\beta_0(\tau))^3$$

for some $\tilde{b}$ that lies between b and $\beta_0(\tau)$. Because of the boundedness $f(0|X_s)$ and its derivatives up to the second order in Assumption 3-i) and the condition in Assumption 4-iii), each term on the right-hand side is finite. Taking expected values on both sides of the equation after multiplying by X_s, and assuming that $\beta_0(\tau) = 0$ without loss of generality, we can obtain that

$$\mathbb{E}[Z_s(b)] - \mathbb{E}[Z_s(\beta_0(\tau))] = \left(\mathbb{E}[X_s1(e_s \leq 0)] - \mathbb{E}[X_s1(e_s \leq X_s'(b-\beta_0))]\right)$$

$$= Ab + Bb^2 + Cb^3,$$

for any $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$, where the coefficients of the polynomial approximation function A, B, and C are given by

$$A = -\mathbb{E}[f(0|X_s)X_s^2], \ B = -\frac{1}{2}\mathbb{E}[f'(0|X_s)X_s^3], \text{ and } C = -\frac{1}{6}\mathbb{E}[f''(X_s\tilde{b}|X_s)X_s^4],$$

respectively. The result then indicates that

$$\sqrt{T}\left(\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||Ab + Bb^2 + Cb^3||\right) = O_p(1), \quad (80)$$

and thus we can obtain that

$$\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T} e_T(s)\right\| \cdot \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left\|\mathbb{E}[Z_s(b)-Z_s(\beta_0(\tau))]\right\|$$

$$\leq \sup_{1\leq s\leq T}|e_T(s)| \cdot \sqrt{T}\left(\sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||Ab + Bb^2 + Cb^3||\right) = O_p\left(\frac{1}{T}\right).$$

This result, together with the $o_p(1)$ terms in (79), leads us to conclude that

$$\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} \left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} e_T(s)(Z_s(b) - Z_s(\beta_0(\tau))) \right\| = o_p(1).$$

For the second term in (78), note that $\mathbb{E}[T^{-1/2} \sum_{s=1}^{T} e_T(s) Z_s(\beta_0(\tau))] = 0$ and its variance is bounded by

$$\begin{aligned} var\left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} e_T(s) Z_s(\beta_0(\tau)) \right) &\leq \sup_{1 \leq s \leq T} |e_T(s)| \cdot var\left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} Z_s(\beta_0(\tau)) \right) \\ &\leq O\left(\frac{1}{T}\right) \cdot \left(\sum_{s=-\infty}^{\infty} ||\mathbb{E}[Z_t Z_{t+s}]|| \right) = O\left(\frac{1}{T}\right), \end{aligned}$$

which implies that the second term in (78) is $o_p(1)$ by Markov inequality. Summing up so far, we showed that

$$||A_1|| \leq \underbrace{\left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \hat{Z}_s \right\|}_{=o_p(1)} \cdot \underbrace{\left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} e_T(s) \hat{Z}_s \right\|}_{=o_p(1)} = o_p(1).$$

The proof for $A_2 = o_p(1)$ can be performed in the same manner. Combining the results of $A_j = o_p(1)$ for $j \in \{1, 2, 3\}$ together, we can infer that

$$\hat{\Omega}_h(\tau) = \hat{\Omega}_h^*(\tau) + o_p(1). \tag{81}$$

Next, we want to show that

$$\hat{\Omega}_h^*(\tau) = \tilde{\Omega}_h^*(\tau) + o_p(1), \tag{82}$$

where $\tilde{\Omega}_h^*(\tau)$ is the infeasible version of $\hat{\Omega}_h^*(\tau)$, i.e.,

$$\tilde{\Omega}_h^*(\tau) = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^*\left(\frac{t}{T}, \frac{s}{T}\right) Z_t Z_s'.$$

To prove this, we define

$$\hat{\Omega}_h^*(\tau; b) := \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^*\left(\frac{t}{T}, \frac{s}{T}\right) Z_t(b) Z_s'(b),$$

for $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$. Using (15), we can re-express $\hat{\Omega}_h^*(\tau; b)$ as:

$$\hat{\Omega}_h^*(\tau; b) = \sum_{k=1}^{\infty} \lambda_k \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) Z_t(b) \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Phi_k\left(\frac{s}{T}\right) Z_s(b) \right)'.$$

Given $k \in \mathbb{N}$, we define that

$$R_{k,T}(b) := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k\left(\frac{t}{T}\right) (Z_t(b) - Z_t).$$

By triangle inequality,

$$\begin{aligned}\|R_{k,T}(b)\| &= \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) \left(X_t \{ 1(e_t \leq X_i'(b - \beta_0(\tau))) - 1(e_t \leq 0) \} \right) \right\| \\ &\leq \sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\check{\mathbb{G}}_{k,T}(b) - \check{\mathbb{G}}_{k,T}(\beta_0(\tau))|| + \sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\tilde{\mathbb{G}}_{k,T}(b) - \tilde{\mathbb{G}}_{k,T}(\beta_0(\tau))|| \\ &\quad + \sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\Lambda_k(b) - \Lambda_k(\beta_0(\tau))||,\end{aligned}$$

where

$$\Lambda_k(b) := \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) \right) \cdot \mathbb{E}[Z_t(b)].$$

The results in Lemma 4 indicate that

$$\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\check{\mathbb{G}}_{k,T}(b) - \check{\mathbb{G}}_{k,T}(\beta_0(\tau))|| = o_p(1) \text{ and } \sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\tilde{\mathbb{G}}_{k,T}(b) - \tilde{\mathbb{G}}_{k,T}(\beta_0(\tau))|| = o_p(1)$$

hold uniformly over $k \in \mathbb{N}$. Also, in view of (80), we have

$$\begin{aligned}\sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||\Lambda_k(b) - \Lambda_k(\beta_0(\tau))|| &= \left\| \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) \right) \cdot \left(\mathbb{E}[X_t 1(e_t \leq X_t'(b - \beta_0(\tau)))] - \mathbb{E}[X_t 1(e_t \leq 0)] \right) \right\| \\ &= \left\| \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) \right) \cdot \left(Ab + Bb^2 + Cb^3 \right) \right\| \\ &\leq \underbrace{\left\| \frac{1}{T} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) \right\|}_{=o(1)} \cdot \underbrace{\left(\sqrt{T} \sup_{b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))} ||Ab + Bb^2 + Cb^3|| \right)}_{=O_p(1)} = o_p(1).\end{aligned}$$

Combining the results together, we can conclude that $||R_{k,T}(b)|| = o_p(1)$ holds uniformly over $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$ and $k \in \mathbb{N}$. Additionally, we have that $T^{-1/2} \sum_{t=1}^{T} \Phi_k(t/T) Z_t = O_p(1)$ uniformly over $k \in \mathbb{N}$. The result, together with $||R_{k,T}(b)|| = o_p(1)$, implies that $T^{-1/2} \sum_{t=1}^{T} \Phi_k(t/T) Z_t(b) = O_p(1)$ uniformly over $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$ and $k \in \mathbb{N}$. Therefore, we can express $\hat{\Omega}_h^*(\tau; b)$ as:

$$\begin{aligned}\hat{\Omega}_h^*(\tau; b) &= \sum_{k=1}^{\infty} \lambda_k \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) Z_t + R_{k,T}(b) \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Phi_k \left(\frac{s}{T} \right) Z_s + R_{k,T}(b) \right)' \\ &= \sum_{k=1}^{\infty} \lambda_k \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) Z_t \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Phi_k \left(\frac{s}{T} \right) Z_s \right)' \\ &\quad + \sum_{k=1}^{\infty} \lambda_k R_{k,T}(b) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Phi_k \left(\frac{s}{T} \right) Z_s \right)' + \sum_{k=1}^{\infty} \lambda_k \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) Z_t \right) R_{k,T}(b)' + \sum_{k=1}^{\infty} \lambda_k R_{k,T}(b) R_{k,T}(b)' \\ &= \underbrace{\sum_{k=1}^{\infty} \lambda_k \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_k \left(\frac{t}{T} \right) Z_t \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Phi_k \left(\frac{s}{T} \right) Z_s \right)'}_{:=\tilde{\Omega}_h^*(\tau)} + o_p(1). \qquad (83)\end{aligned}$$

The last equation holds because $R_{k,T}(b)$ shrinks to zero uniformly over $b \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$ and $k \in \mathbb{N}$, and

$$\begin{aligned}\left\|\sum_{k=1}^{\infty}\lambda_k R_{k,T}(b)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Phi_k\left(\frac{s}{T}\right)Z_s\right)'\right\| &\leq \left(\sup_{k\in\mathbb{N}}||R_{k,T}(b)||\right)\left(\sum_{k=1}^{\infty}\lambda_k\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Phi_k\left(\frac{s}{T}\right)Z_s\right\|\right)\\ &= o_p(1), \qquad (84)\end{aligned}$$

where (84) follows from $\sup_{k\in\mathbb{N}}||R_{k,T}(b)|| = o_p(1)$, $\sum_{k=1}^{\infty}\lambda_k = O(1)$, and the fact that $var(T^{-1/2}\sum_{s=1}^{T}\Phi_k(s/T)Z_s)$ is bounded uniformly over k. Additionally, we have that

$$\left\|\sum_{k=1}^{\infty}\lambda_k R_{k,T}(b)R_{k,T}(b)'\right\| \leq \left(\sup_{k\in\mathbb{N},\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}||R_{k,T}(b)||\right)^2 \cdot \left(\sum_{k=1}^{\infty}\lambda_k\right) = o_p(1),$$

where the equation follows from $\sum_{k=1}^{\infty}\lambda_k = \int_0^1 Q_h(r,r)dr = O(1)$, and $\sup_{k\in\mathbb{N},\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}||R_{k,T}(b)|| = o_p(1)$. From this finding and $\hat{\beta}(\tau) \in \mathcal{N}_{\epsilon_T}(\beta_0(\tau))$, with probability arbitrarily close to one, we can conclude that

$$||\hat{\Omega}_h^*(\tau) - \tilde{\Omega}_h^*(\tau)|| \leq \sup_{b\in\mathcal{N}_{\epsilon_T}(\beta_0(\tau))}\left\|\hat{\Omega}_h^*(\tau;b) - \tilde{\Omega}_h^*(\tau)\right\| + o_p(1).$$

Using the same reasoning as shown for $\hat{\Omega}_h(\tau) = \hat{\Omega}_h^*(\tau) + o_p(1)$, we can show that $\tilde{\Omega}_h^*(\tau) = \tilde{\Omega}_h(\tau) + o_p(1)$ holds for all h, where

$$\tilde{\Omega}_h(\tau) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_{T,h}^*\left(\frac{t}{T},\frac{s}{T}\right)Z_tZ_s' = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_h\left(\frac{t}{T},\frac{s}{T}\right)Z_t^cZ_s^{c\prime}$$

with $Z_t^c = Z_t - T^{-1}\sum_{s=1}^{T}Z_s$. Combining the result into $\hat{\Omega}_h(\tau) = \hat{\Omega}_h^*(\tau) + o_p(1)$ and (82), we can conclude that

$$\begin{aligned}\hat{\Omega}_h(\tau) - \tilde{\Omega}_h(\tau) &= (\hat{\Omega}_h(\tau) - \hat{\Omega}_h^*(\tau)) + (\hat{\Omega}_h^*(\tau) - \tilde{\Omega}_h^*(\tau)) + (\tilde{\Omega}_h^*(\tau) - \tilde{\Omega}_h(\tau))\\ &= o_p(1)\end{aligned}$$

holds for any fixed h, which is the desired result. For the result in (23), we use

$$\hat{\Omega}_h(\tau) = \hat{\Omega}_h^*(\tau) + o_p(1) = \tilde{\Omega}_h^*(\tau) + o_p(1),$$

where $\tilde{\Omega}_h^*(\tau)$ is defined in (83). Then, (23) follows directly from the result in Lemma 6. ■

Proof of Theorem 2. Let $\Delta(\tau)$ be a $p \times p$ matrix such that $\Delta(\tau)\Delta(\tau)' = RD(\tau)^{-1}\Omega(\tau)D(\tau)^{-1}R'$. From (4) and the result in Theorem 1, and by applying Slutsky's theorem, we obtain

$$\begin{aligned}&\sqrt{T}R(\hat{\beta}(\tau) - \beta_0(\tau)) \xrightarrow{d} \Delta(\tau)\mathbb{Z}_p;\\ &R\hat{\Sigma}(\tau)R' \xrightarrow{d} \underbrace{(RD(\tau)^{-1}\Omega(\tau)^{1/2})\mathbb{S}_{h,\infty}(\Omega(\tau)^{1/2\prime}D(\tau)^{-1}R')}_{\stackrel{d}{=}\Delta(\tau)\mathbb{S}_{h,\infty}^{[p]}\Delta(\tau)'},\end{aligned}$$

where the weak convergences hold jointly. Note that $\mathbb{Z}_p = W_p(1)$ and $\mathbb{S}_{h,\infty}^{[p]} = \sum_{j=1}^{\infty} \lambda_j \mathbb{Z}_{p,j} \mathbb{Z}_{p,j}'$ with $\mathbb{Z}_{p,j} := \int_0^1 \Phi_j(r)\, dW_p(r) \overset{\text{i.i.d.}}{\sim} N(0, I_p)$. $\mathbb{Z}_p$ and $\mathbb{S}_{h,\infty}^{[p]}$ are independent because

$$
\begin{aligned}
Cov(\mathbb{Z}_p, \mathbb{Z}_{p,j}) &= Cov\left(W_p(1), \int_0^1 \Phi_j(r)\, dW_p(r)\right) \\
&= Cov\left(\int_0^1 dW_p(r), \int_0^1 \Phi_j(r)\, dW_p(r)\right) \\
&= I_p \cdot \int_0^1 \Phi_j(r)\, dr
\end{aligned}
$$

for all $j \in \mathbb{N}$. This allows us to obtain the result in part (a) using the continuous mapping theorem. The result with $p = 1$ can be shown in a similar manner. For the result in part (b), it follows directly from the fact that $\mathbb{S}_{h,\infty}^{[p]}$ can also be represented as $K^{-1}\mathbb{W}_p(K, I_p)$, and from the equivalence between Hotelling's T-squared random variable and a scaled F-random variable, as discussed in Sun (2013, pp. 6). ■

References

[1] Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. American Economic Review, 109(4), 1263-1289.

[2] Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica: Journal of the Econometric Society, 817-858.

[3] Angrist, J., Chernozhukov, V., & Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the US wage structure. Econometrica, 74(2), 539-563.

[4] Bofingeb, E. (1975). Estimation of a density function using order statistics 1. Australian journal of statistics, 17(1), 1-7.

[5] Chen, X., Liao, Z., & Sun, Y. (2014). Sieve inference on possibly misspecified semi-nonparametric time series models. Journal of Econometrics, 178, 639-658.

[6] de Castro, L., Galvao, A. F., Kaplan, D. M., & Liu, X. (2019). Smoothed GMM for quantile models. Journal of Econometrics, 213(1), 121-144.

[7] Fitzenberger, B. (1998). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. Journal of Econometrics, 82(2), 235-287.

[8] Galvao, A. F., & Kato, K. (2016). Smoothed quantile regression for panel data. Journal of econometrics, 193(1), 92-112.

[9] Galvao, A. F., & Yoon, J. (2024). Hac covariance matrix estimation in quantile regression. Journal of the American Statistical Association, 119(547), 2305-2316.

[10] Gregory, K. B., Lahiri, S. N., & Nordman, D. J. (2018). A smooth block bootstrap for quantile regression with time series. The Annals of Statistics, 46(3), 1138-1166.

[11] Hannan, E. J., Terrell, R. D., & Tuckwell, N. E. (1970). The seasonal adjustment of economic time series. International Economic Review, 11(1), 24-52.

[12] Hoga, Yannick, and Christian Schulz. "Self-Normalized Inference in (Quantile, Expected Shortfall) Regressions for Time Series." arXiv preprint arXiv:2502.10065 (2025).

[13] Hwang, J., & Sun, Y. (2017). Asymptotic F and t tests in an efficient GMM setting. Journal of econometrics, 198(2), 277-295.

[14] Hwang, J., & Sun, Y. (2018). Should we go one step further? An accurate comparison of one-step and two-step procedures in a generalized method of moments framework. Journal of econometrics, 207(2), 381-405.

[15] Hwang, J., & Valdés, G. (2023). Finite-sample corrected inference for two-step GMM in time series. Journal of Econometrics, 234(1), 327-352.

[16] Jenish, N., & Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of econometrics, 150(1), 86-98.

[17] Kiefer, N. M., & Vogelsang, T. J. (2002). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. Econometrica, 70(5), 2093-2095.

[18] Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory, 21(6), 1130-1164.

[19] Knessl, C., & Keller, J. B. (1991). Asymptotic properties of eigenvalues of integral equations. SIAM Journal on Applied Mathematics, 51(1), 214-232.

[20] Koenker, R. (2005). Quantile regression (Vol. 38). Cambridge university press.

[21] Lazarus, E., Lewis, D. J., & Stock, J. H. (2021). The Size-Power Tradeoff in HAR Inference. Econometrica, 89(5), 2497-2516.

[22] Lazarus, E., Lewis, D. J., Stock, J. H., & Watson, M. W. (2018). HAR inference: Recommendations for practice. Journal of Business & Economic Statistics, 36(4), 541-559.

[23] Lee, S., Liao, Y., Seo, M. H., & Shin, Y. (2024). Fast inference for quantile regression with tens of millions of observations. Journal of Econometrics, 105673.

[24] Merlevède, F., Peligrad, M., & Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. Probability Theory and Related Fields, 151, 435-474.

[25] Müller, U. K. (2007). A theory of robust long-run variance estimation. Journal of Econometrics, 141(2), 1331-1352.

[26] Müller, U. K. (2014). HAC corrections for strongly autocorrelated time series. Journal of Business & Economic Statistics, 32(3), 311-322.

[27] Müller, U. K., & Watson, M. W. (2020). Low-frequency analysis of economic time series. preparation. In Handbook of Econometrics. Elsevier.

[28] Newey, W. K. & West, K. D. (1987), A simple, positive semi-de nite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

[29] Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. The Review of Economic Studies, 61(4), 631-653.

[30] Pakes, A., & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. Econometrica: Journal of the Econometric Society, 1027-1057.

[31] Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. The Annals of Mathematical Statistics, 329-348.

[32] Phillips, P. C. (2005). HAC estimation by automated regression. Econometric Theory, 21(1), 116-142.

[33] Phillips, P. C., Sun, Y., & Jin, S. (2006). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. International Economic Review, 47(3), 837-894.

[34] Hall, P., & Sheather, S. J. (1988). On the distribution of a studentized quantile. Journal of the Royal Statistical Society: Series B (Methodological), 50(3), 381-391.

[35] Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. Journal of the Royal Statistical Society Series B: Statistical Methodology, 72(3), 343-366.

[36] Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. Journal of Econometrics, 164(2), 345-366.

[37] Sun, Y. (2013). A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. The Econometrics Journal, 16(1), 1-26.

[38] Sun, Y. (2014a). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. Journal of Econometrics, 178, 659-677.

[39] Sun, Y. (2014b). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. Econometrica, 82(6), 2327-2370.

[40] Sun, Y., & Phillips, P. C. (2008). Optimal bandwidth choice for interval estimation in GMM regression.

[41] Sun, Y., Phillips, P. C., & Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. Econometrica, 76(1), 175-194.

[42] Tuan, D., & Lanh, T. (1985). Some mixing properties of time series models. Stochastic processes and their applications, 19, 297-303.

[43] Van Der Vaart, A. W., Wellner, J. A., van der Vaart, A. W., & Wellner, J. A. (1996). Weak convergence (pp. 16-28). Springer New York.

[44] Xiao, Z. (2009). Quantile cointegrating regression. Journal of econometrics, 150(2), 248-260.

[45] Xu, K. L. (2021). On the serial correlation in multi-horizon predictive quantile regression. Economics Letters, 200, 109736.

[46] Ye, X., & Sun, Y. (2018). Heteroskedasticity-and autocorrelation-robust F and t tests in Stata. The Stata Journal, 18(4), 951-980.

[47] White, H., & Domowitz, I. (1984). Nonlinear regression with dependent observations. Econometrica: Journal of the Econometric Society, 143-161.

[48] Xiao, Z. (2012). Time series quantile regressions. In Handbook of statistics (Vol. 30, pp. 213-257). Elsevier.

[49] Yoshihara, K. I. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 35(3), 237-252.

[50] Zhou, Z., & Shao, X. (2013). Inference for linear models with dependent errors. Journal of the Royal Statistical Society Series B: Statistical Methodology, 75(2), 323-343.