Testing Structural Change in Conditional Distributions via Quantile Regressions *

Liangjun Su,[a] Zhijie Xiao[b]

[a]School of Economics, Singapore Management University, Singapore

[b]Department of Economics, Boston College, Chestnut Hill, MA, USA

September 22, 2009

Abstract

We propose tests for structural change in conditional distributions via quantile regressions. To avoid misspecification on the conditioning relationship, we construct the tests based on the residuals from local polynomial quantile regressions. In particular, the tests are based upon the cumulative sums of generalized residuals from quantile regressions and have power against local alternatives at rate $n^{-1/2}$. We derive the limiting distributions for our tests under the null hypothesis of no structural change and a sequence of local alternatives. The proposed tests apply to a wide range of dynamic models, including time series regressions with m.d.s. errors, as well as models with serially correlated errors. To deal with possible correlations in the error process, we also propose a simulation method to obtain the p-values for our tests. Finally, Monte Carlo simulations suggest that our tests behave well in finite samples.

JEL classifications: C12, C14, C22, C5

Key Words: Conditional distribution, Structural change, Local polynomial regression, Quantile regression, Block bootstrap

*Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; Phone: +65 6828 0386; e-mail: ljsu@smu.edu.sg. Zhijie Xiao, Department of Economics, Boston College, Chestnut Hill, MA 02467, USA; phone: +1 617 5521709; e-mail: xiaoz@bc.edu. The first author gratefully acknowledges the financial support from a research grant (Grant number: C244/MSS8E004) from Singapore Management Unviersity.

1 Introduction

Structural instability is an empirically widespread phenomenon. The presence of structural instability may invalidate the conventional econometric inference that ignores it. For this reason, there has traditionally been a long-standing research effort in structural instability, among which much has been devoted to parametric models, especially linear models. In particular, there has been a large literature on testing parameter instability in linear regression models. See, inter alia, Page (1955), Brown, Durbin, and Evans (1975), Nyblom (1989), Ghysels and Hall (1990), Andrews (1993), Andrews and Ploberger (1994), Sowell (1996), Bai, Lumsdaine and Stock (1998), Hansen (2000), Elliott and Müller (2006), and Li (2008). See Csörgő and Horváth (1997) for an excellent review on this topic.

An important subfield that has attracted a lot of research attention is testing for structural change at the distributional level. Many economic and finance problems often confront structural instability in distribution. Demographic changes, capital accumulation, technological progress, and many other aspects in economic environment and policies, often change the distribution or conditional distribution of economic time series. Hansen (2000) noted that tests for structural changes in regression models are sensitive to a change in the marginal distribution of the regressors. Qu (2008) argued that structural change in conditional distributions may be of key importance under various circumstances. Take the income inequality study as an example. It is important to examine whether the distribution of wage differentials between different races or genders, controlled for relevant covariates, has changed over time or not. It can be the case where the conditional mean of the wage differentials remain unchanged but the conditional dispersion has changed. If so, traditional tests for structural changes that are based on the conditional mean regression models should be replaced by tests for conditional dispersion or distribution. In other scenarios, higher order conditional moments may have changed over time but both the conditional mean and dispersion remain unchanged. If this is the case, then the conditional mean and variance cease to be informative and tests in the conditional distributions should be employed. In light of this, it is important to test the distributional stability of a time series in regression models.

Several tests have been developed to test for distributional changes. Picard (1985) proposed tests for distributional change in time series by detecting changes in the spectrum in that time series. Bai (1994) considered tests for the distributional change in the i.i.d. error process of ARMA models based upon empirical distributions. Inoue (2001) proposed tests for distributional change based on a sequential empirical process for dependent data. Lee and Na (2004) proposed tests for distributional change based upon nonparametric kernel density estimation in the time series framework. All these tests are designed to test for structural changes in *unconditional* distributions.

In this paper, we study testing for structural change in the *conditional* distribution of a random variable Y_t given relevant covariates X_t, where X_t may include lagged variables of Y_t. We propose tests for distributional changes based on quantile regressions. Being the inverse of a conditional distribution function, the conditional quantile function is a natural object to examining conditional distributional changes. In the special case where the relationship between Y_t and X_t is characterized by a parametric model, testing for distributional change may be formulated as testing quantile regression coefficient instability. Su and Xiao (2008a) and Qu (2008) proposed tests for parameter instability in linear quantile regression models. These tests can be applied to test for structural changes in conditional distribution if the specified linear relationship between Y_t and X_t is correct. In many applications, the functional form of the relationship between Y_t and X_t is unknown. Misspecification of econometric models can also manifest themselves in the form of structural changes. Misleading conclusions may be obtained if the linearity (or other parametric) assumption is violated.

To avoid spurious breaks from misspecification, we propose tests for distributional changes via nonparametric quantile regressions. Chaudhuri (1991) studied nonparametric quantile regression in the i.i.d. setting and derived its local Bahadur representation. Su and White (2009b) studied time series local polynomial quantile regression and established the uniform local Bahadur representation, where the uniformity holds in both quantiles and conditioning variables. Also see Yu and Jones (1998), Koenker, Ng, and Portnoy (1994) for other studies in nonparametric quantile regressions. In this paper, we use local polynomial quantile regressions to construct the proposed tests.

There are several important features associated with our tests. First, our tests are testing for structural changes at the distributional level without specifying any parametric form on any aspect of the conditional distribution, including conditional mean, conditional variance, or conditional quantile function. Second, comparing our tests with the existing literature, we consider tests for structural changes in the *conditional* distribution for *time series* data. This is important since economic and financial time series are not i.i.d. and conditional distributions may be affected by policy changes or critical social events. Third, our tests are flexible on the model dynamics and do not require the correct specification of the dynamics. Letting $\mathcal{F}_{t-1}$ be the information set at time t, we do not require that the distribution of Y_t conditional on $\mathcal{F}_{t-1}$ be the same as that of Y_t conditional on X_t. As a result, our tests cover a wide range of dynamic models and do not require that the error process in the quantile regression model be a martingale difference sequence (m.d.s. hereafter). As we will demonstrate, our tests are asymptotically pivotal if the m.d.s. condition is satisfied. For more general case, we propose a simulation method to facilitate statistical inference. Fourth, as in Su and White (2009a) we allow for small breaks in the covariate process $\{X_t\}$ under both the null and alternative hypotheses. Fifth, our tests allow us to focus on certain range of the conditional distributions. For example, one may focus on the median or (say) left-tail of the conditional distributions as in the value-at-risk (VaR) analysis. Finally, even though our tests are of nonparametric nature, they have non-trivial power against a sequence of Pitman local alternatives that converge to zero at the parametric $n^{-1/2}$-rate.

The rest of the paper is organized as follows. In Section 2 we introduce our hypotheses, local polynomial quantile regression estimates and test statistics. In Section 3 we study the asymptotic properties of our test statistics and propose a method to simulate the p-values. In Section 4 we provide a small set of Monte Carlo experiments to evaluate the finite sample performance of our tests. Section 5 contains concluding remarks. All proofs are relegated to the appendix.

A word on notation. Throughout the paper, we use $\mathbf{1}(\cdot)$ to denote the indicator function and $\|\cdot\|$ to denote the Euclidean norm. Let $\pi_1 \wedge \pi_2 \equiv \min(\pi_1, \pi_2)$, where $x \equiv y$ indicates that x is defined by y or y is defined by x, which is clear from the context. The operators $\xrightarrow{P}$ and $\xrightarrow{D}$ denote convergence in probability and distribution, respectively. We use $\Rightarrow$ to denote weak convergence and $\overset{p}{\Rightarrow}$ to denote weak convergence in probability as defined by Giné and Zinn (1990); see also Hansen (2000).

2 Hypotheses and Tests

2.1 Hypotheses

Let $\{(Y_{nt}, X_{nt})\}_{t=1}^{n}$ be a sequence of time series random vectors, we are interested in testing the null hypothesis of no change in the conditional distribution of Y_{nt} given $X_{nt} \in \mathbb{R}^d$. The triangular-array notation $\{(Y_{nt}, X_{nt})\}_{t=1}^{n}$ facilitates the study of asymptotic local power properties of our test and allows for small deviations from stationarity. To avoid complicated notation we will mostly suppress reference to the n subscript in what follows, in particular, we write $Y_t = Y_{nt}$, $X_t = X_{nt}$. If we denote the conditional distribution function of Y_t given X_t as $F_t(\cdot|X_t)$, the null hypothesis can be written as

$$H_0^* : F_t(\cdot|X_t) = F_0(\cdot|X_t) \text{ a.s. for some } F_0(\cdot|\cdot) \text{ and all } t = 1, 2, \cdots \tag{2.1}$$

Alternatively, since the conditional quantile function is the inverse function of the conditional distribution function, we may also equivalently write the null hypothesis as

$$H_0 : F_t^{-1}(\tau, X_t) = F_0^{-1}(\tau, X_t) \text{ a.s. for some } F_0^{-1}(\cdot, \cdot) \text{ and all } t = 1, 2, \cdots \tag{2.2}$$

where $F_t^{-1}(\tau, x)$ is the τth conditional quantile function of Y_t given $X_t = x$, that is,

$$F_t^{-1}(\tau, x) \equiv \inf \{y : F_t(y|x) \geq \tau\}.$$

To test for structural changes in the conditional distribution of Y_t, we may construct appropriate estimation for the conditional quantile function or the conditional distribution function and examine their stability over time. In this paper, we propose testing procedures for distributional changes via quantile regressions.

For notational convenience, we denote $F_t^{-1}(\tau, x) = m_t(\tau, x)$, and $F_0^{-1}(\tau, x) = m_0(\tau, x)$, so that the null hypothesis can be written as

$$H_0 : m_t(\tau, X_t) = m_0(\tau, X_t) \text{ a.s. for some } m_0(\cdot, \cdot) \text{ and all } t = 1, 2, \cdots, n. \tag{2.3}$$

The alternative hypothesis is the negation of H_0^*. In this paper, we study the asymptotic behavior of the proposed test under a sequence of Pitman local alternatives:

$$H_{1n} : m_t(\tau, X_t) = m_0(\tau, X_t) + n^{-1/2}\delta(\tau, X_t, t/n), \tag{2.4}$$

where $\delta(\cdot, \cdot, \cdot)$ is a non-constant measurable function. If $\delta(\tau, x, t/n) = \delta_0(\tau, x)\mathbf{1}(t/n \geq \pi_0)$ in eq. (2.4), we have the special case of a one-time shift at time $n\pi_0$.

In practice, the functional form of the conditional distribution is usually unknown and misspecification of the conditional relationship manifests themselves in the form of structural changes. For this reason, we propose tests for distribution changes based on nonparametric quantile regressions.

2.2 Estimation

The approach proposed in this paper may be applied to different nonparametric estimators, including the simple kernel smoother and the local polynomial estimator. In this paper, we give asymptotic analysis based on the local polynomial procedures. The basic idea of local polynomial fit is: if $m_0(\tau, x)$ is a smooth function of x, for any x_i in a neighborhood of x, we have

$$\begin{aligned} m_0(\tau, x_i) &\simeq m_0(\tau, x) + \sum_{1\le|\mathbf{j}|\le p} \frac{1}{\mathbf{j}!} D^{|\mathbf{j}|} m_0(\tau, x)(x_i - x)^{\mathbf{j}} + o(\|x_i - x\|^p) \\ &\equiv \sum_{0\le|\mathbf{j}|\le p} \beta_{\mathbf{j}}(\tau, x; h)((x_i - x)/h)^{\mathbf{j}} + o(\|x_i - x\|^p). \end{aligned}$$

Here, we use the notation of Masry (1996): $\mathbf{j} = (j_1, ..., j_d)$, $|\mathbf{j}| = \sum_{i=1}^d j_i$, $x^{\mathbf{j}} = \mathbf{\Pi}_{i=1}^d x_i^{j_i}$, $\sum_{0\le|\mathbf{j}|\le p} = \sum_{k=0}^p \sum_{\substack{j_1=0 \\ j_1+...+j_d=k}}^k \cdots \sum_{j_d=0}^k$, $D^{|\mathbf{j}|} m_0(\tau, x) = \frac{\partial^{|\mathbf{j}|} m(\tau,x)}{\partial^{j_1} x_1 ... \partial^{j_d} x_d}$, $\beta_{\mathbf{j}}(\tau, x; h) = \frac{h^{|\mathbf{j}|}}{\mathbf{j}!} D^{|\mathbf{j}|} m_0(\tau, x)$, where $\mathbf{j}! \equiv \mathbf{\Pi}_{i=1}^{d_1} j_i!$ and $h = h(n)$ is a bandwidth parameter that controls how "close" x_i is from x. Thus, given observations $\{(Y_t, X_t)\}_{t=1}^n$, we may consider a local-polynomial quantile regression that minimizes the following objective function

$$Q_n(\tau, x; \boldsymbol{\beta}) \equiv n^{-1} \sum_{t=1}^n \rho_\tau \left(Y_t - \sum_{0\le|\mathbf{j}|\le p} \beta_{\mathbf{j}}((X_t - x)/h)^{\mathbf{j}} \right) K((x - X_t)/h), \tag{2.5}$$

where $\rho_\tau(z)$ be the "check" function defined by $\rho_\tau(z) = z(\tau - \mathbf{1}(z \le 0))$, K is a nonnegative kernel function on $\mathbb{R}^d$, and $\boldsymbol{\beta}$ is a stack of $\beta_{\mathbf{j}}$ in the lexicographical order (with highest priority to last position so that $(0, 0, ..., l)$ is the first element in the sequence and $(l, 0, ..., 0)$ is the last element). Minimizing (2.5) with respect to $\beta_{\mathbf{j}}$, $0 \le |\mathbf{j}| \le p$, delivers an estimate $\widehat{\beta}_{\mathbf{j}}(\tau, x; h)$ of $\beta_{\mathbf{j}}(\tau, x; h)$. The conditional quantile function $m_0(\tau, x)$ and its derivatives up to p-th order are then estimated respectively by

$$\widehat{m}(\tau, x) = \widehat{\beta}_{\mathbf{0}}(\tau, x; h) \text{ and } \widehat{D}^{|\mathbf{j}|} m(\tau, x) = (\mathbf{j}!/h^{|\mathbf{j}|}) \widehat{\beta}_{\mathbf{j}}(\tau, x; h) \text{ for } 1 \le |\mathbf{j}| \le p.$$

In the special case when $p = 1$, these are the widely used local linear estimators.

Let $N_l = (l + d - 1)!/(l!(d - 1)!)$ be the number of distinct d-tuples $\mathbf{j}$ with $|\mathbf{j}| = l$. It denotes the number of distinct l-th order partial derivatives of $m_0(\tau, x)$ with respect to x. Arrange the N_l d-tuples as a sequence in the lexicographical order, and let ϕ_l^{-1} denote this one-to-one map. For each $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2p$, let $\mu_{\mathbf{j}} = \int_{\mathbb{R}^d} x^{\mathbf{j}} K(x) dx$, and define the $N \times N$ dimensional matrix $\mathbb{H}$ and $N \times 1$ vector $\mathbb{B}$, where $N = \sum_{l=1}^p N_l$, by

$$\mathbb{H} = \begin{bmatrix} \mathbb{H}_{0,0} & \mathbb{H}_{0,1} & \cdots & \mathbb{H}_{0,p} \\ \mathbb{H}_{1,0} & \mathbb{H}_{1,1} & \cdots & \mathbb{H}_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{H}_{p,0} & \mathbb{H}_{p,1} & \cdots & \mathbb{H}_{p,p} \end{bmatrix}, \ \mathbb{B} = \begin{bmatrix} \mathbb{H}_{0,0} \\ \mathbb{H}_{1,0} \\ \vdots \\ \mathbb{H}_{p,0} \end{bmatrix}, \tag{2.6}$$

where $\mathbb{H}_{i,j}$ are $N_i \times N_j$ dimensional matrices whose (l, r) elements are $\mu_{\phi_i(l)+\phi_j(r)}$. That is, the elements of $\mathbb{H}$ and $\mathbb{B}$ are simply multivariate moments of the kernel K. In addition, we denote $c_0 = e_1' \mathbb{H}^{-1} \mathbb{B}$. One can verify that $c_0 = 1$ if $p = 1$ and a symmetric kernel function $K(\cdot)$ is applied, and c_0 lies strictly between 0 and 1 for general local polynomial regression with $p \ge 2$.

For additional information about local polynomial estimation, see Fan (1992), and Fan and Gijbels (1996) for discussions on the attractive properties of this approach, and Chaudhuri (1991), Fan, Hu, and Truong (1994), Yu and Jones (1998), and Su and White (2009b) for studies on local-polynomial quantile regressions.

Let $\mu((X_t - x)/h)$ be an $N \times 1$ vector that contains the regressors $((X_t - x)/h)^{\mathbf{j}}$ in the local-polynomial quantile regression (2.5) in the lexicographical order. For example, if $p = 1$, then $\mu((X_t - x)/h) = (1, (X_t - x)'/h)'$. Define

$$H_n(\tau, x) \equiv \frac{1}{n}\sum_{t=1}^{n} f_t(m_0(\tau, x)|x) f_t(x) \mathbb{H}, \text{ and} \tag{2.7}$$

$$J_n(\tau, x) \equiv \frac{1}{\sqrt{nh^d}}\sum_{t=1}^{n} \psi_\tau(Y_t - m_0(\tau, X_t)) \mu((X_t - x)/h) K((x - X_t)/h). \tag{2.8}$$

The following result is essentially Corollary 2.2 of Su and White (2009b).

Proposition 2.1 *Suppose that H_{1n} and Assumptions A1-A7 given below hold. Then uniformly in $(\tau, x) \in \mathcal{T} \times \mathcal{X}$,*

$$\sqrt{nh^d}(\widehat{m}(\tau, x) - m_0(\tau, x)) = e_1' H_n(\tau, x)^{-1} J_n(\tau, x)[1 + o_P(1)] + o_P(h^{d/2}),$$

where $e_1 = (1, 0, ..., 0)'$ is an N-vector, $\mathcal{T} = [\underline{\tau}, \overline{\tau}] \subset (0, 1)$ and $\mathcal{X}$ is the support of the distribution of X_t.

2.3 The proposed tests

Under the null hypothesis, $F_t(m_0(\tau, X_t)|X_t) = \tau$ a.s. for each t, i.e., $E[\mathbf{1}(Y_t \leq m_0(\tau, X_t))] = \tau$. Let $u_{t\tau} \equiv Y_t - m_0(\tau, X_t)$, and $\psi_\tau(u) \equiv \tau - \mathbf{1}(u \leq 0)$, then

$$E[\psi_\tau(u_{t\tau})] = 0 \text{ under } H_0.$$

This suggests, if $u_{t\tau}$ were observable, one could test H_0 based on the following process

$$S_n^{(1)}(\pi, \tau) = n^{-1/2}\sum_{t=1}^{\lceil n\pi \rfloor} \psi_\tau(u_{t\tau}),$$

where $\lceil c \rfloor$ is the integer part of c. Under the null hypothesis, $\{S_n^{(1)}(\cdot, \cdot)\}$ converges to a zero-mean Gaussian process.

However, $u_{t\tau}$ are not observed in practice. If we replace it by $\widehat{u}_{t\tau} \equiv Y_t - \widehat{m}(\tau, X_t)$, we can consider the following residual-based CUSUM process

$$S_n^{(2)}(\pi, \tau) = n^{-1/2}\sum_{t=1}^{\lceil n\pi \rfloor} \{\tau - \mathbf{1}(\widehat{u}_{t\tau} \leq 0)\}.$$

Note that the indicator function is not everywhere differentiable on its support. Even if we can assume that the conditional quantile function $m_0(\tau, x)$ belongs to certain class of smooth functions (e.g., Van der Vaart and Wellner (1996, p.154)) so that $S_n^{(1)}(\pi, \tau)$ obeys a version of Donsker theorem, it is hard to justify that the estimate $\widehat{m}(\tau, x)$ of $m_0(\tau, x)$ also belongs to the same class. For this reason, we propose to approximate the indicator function by a smooth function $G(\cdot)$ and consider the following two-parameter stochastic process:

$$S_n(\pi,\tau) = n^{-1/2} \sum_{t=1}^{\lceil n\pi \rfloor} \{\tau - G_{\lambda_n}(\widehat{u}_{t\tau})\},$$

where $G_\lambda(u) = G(-u/\lambda)$, and $G(\cdot)$ behaves like a cumulative distribution function (c.d.f. hereafter) and $\lambda_n \to 0$ sufficiently fast as $n \to \infty$. Since $\psi_\tau(u_{t\tau})$ has conditional mean zero given X_t under the null, we can treat elements in the summation of $S_n(\pi,\tau)$ as generalized quantile regression residuals.

The process $\{S_n(\cdot,\cdot), n \geq 1\}$ will be the main ingredient of our test statistics. As we show in the next section, under some regularity conditions, it converges to a zero-mean Gaussian process under the null hypothesis and diverges for some value of (π,τ) under the alternative. However, the estimation $\widehat{m}(\tau,x)$ of $m_0(\tau,x)$ affects the limiting distribution of $\{S_n(\cdot,\cdot), n \geq 1\}$, and thus brings additional difficulty to our inference problem. For this reason, we consider the following centered process

$$S_n^c(\pi,\tau) = S_n(\pi,\tau) - \pi S_n(1,\tau). \tag{2.9}$$

As we will demonstrate later in this paper, re-centering $S_n(\pi,\tau)$ by the quantity $\pi S_n(1,\tau)$ eliminates the preliminary estimation error under some regularity conditions.

Inference procedures may be constructed based on different continuous functionals of $S_n^c(\pi,\tau)$. We consider the leading cases of Kolmogorov-Smirnoff and Cramér-von Mises testing statistics defined as follows

$$KS_n \equiv \sup_{\pi\in[0,1]} \sup_{\tau\in\mathcal{T}} |S_n^c(\pi,\tau)| = \max_{1\leq j\leq n} \sup_{\tau\in\mathcal{T}} |S_n^c(j/n,\tau)|,$$

$$CM_n \equiv \int_{\mathcal{T}} \int_0^1 S_n^c(\pi,\tau)^2 \, d\pi w(d\tau) = \int_{\mathcal{T}} \frac{1}{n} \sum_{j=1}^{n} S_n^c(j/n,\tau)^2 \, w(d\tau),$$

where $\mathcal{T} = [\underline{\tau}, \overline{\tau}]$ is a subset of $(0,1)$, and $w(\tau) = 1/(\overline{\tau} - \underline{\tau})$ if $\tau \in \mathcal{T}$ and 0 otherwise. Of course, other types of integrating functions for τ are possible. We explore the asymptotic properties of the proposed tests in the next section.

3 Asymptotic Theory

3.1 Assumptions

For asymptotic analysis, we make the following assumptions.

Assumption A1. $\{(Y_t, X_t)\} \equiv \{(Y_{nt}, X_{nt})\}$ is a strong mixing process with mixing coefficients $\alpha(s)$ such that $\sum_{s=0}^{\infty} s^5 \alpha(s)^{\eta/(6+\eta)} \leq C < \infty$ for some $\eta > 0$ with $\eta/(6+\eta) \leq 1/2$.

Assumption A2. (i) The probability density function (p.d.f.) $f_t(\cdot) \equiv f_{nt}(\cdot)$ of X_t is bounded with compact support $\mathcal{X}$ and has uniformly bounded first order partial derivatives for each t. (ii) The conditional c.d.f. $F_t(\cdot|X_t) \equiv F_{nt}(\cdot|X_t)$ of Y_t given X_t has Lebesgue density $f_t(\cdot|X_t) \equiv f_{nt}(\cdot|X_t)$ such that $\sup_{n\geq 1} \sup_{y:\, F_t(y|X_t)\in\mathcal{T}} f_t(y|X_t) \leq C_1$ a.s. for all t, and $|f_t(y_1|X_t) - f_t(y_2|X_t)| \leq C_2(X_t)\,|y_1 -$

$y_2|$ a.s. for all t, where $C_2(\cdot)$ is a continuous function. $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} f_i(m_0(\tau,x)|x) f_i(x) > 0$ uniformly in $(\tau,x) \in \mathcal{T}\times\mathcal{X}$. (iii) Let $\mathcal{F}_{t-1} \equiv \sigma(X_{t-i}, i \geq 0; Y_{t-j}, j \geq 1)$. The conditional c.d.f. $F_t(\cdot|\mathcal{F}_{t-1})$ of Y_t given $\mathcal{F}_{t-1}$ has Lebesgue p.d.f. $f_t(\cdot|\mathcal{F}_{t-1})$ that have continuous derivatives up to qth order for $q \geq 2$. The qth derivative $f_t^{(q)}(\cdot|\mathcal{F}_{t-1})$ is uniformly continuous and $\sup_{y:F_t(y|\mathcal{F}_{t-1})\in\mathcal{T}} |f_t^{(q)}(\cdot|\mathcal{F}_{t-1})| < \infty$ a.s. (iv) Let $W_t \equiv (Y_t, X_t')'$. The joint p.d.f. of $(W_{t_1}, W_{t_2}, \cdots, W_{t_{12}})$ exists and is bounded.

Assumption A3. (i) $m_0(\tau,x)$ is bounded uniformly in $(\tau,x)\in\mathcal{T}\times\mathcal{X}$. It is Lipschitz continuous in (τ,x) and has derivatives with respect to x up to order $p+1$. (ii) The $(p+1)$th order partial derivatives with respect to x, i.e., $D^{\mathbf{k}}m_0(\tau,x)$ with $|\mathbf{k}| = p+1$, are uniformly bounded in $(\tau,x)\in\mathcal{T}\times\mathcal{X}$, and are Hölder continuous in (τ,x) with exponent $\gamma_0 > 0$: $||D^{\mathbf{k}}m_0(\tau,x) - D^{\mathbf{k}}m_0(\tau',x')|| \leq C_3\left(|\tau-\tau'|^{\gamma_0} + ||x-x'||^{\gamma_0}\right)$ for some constant $C_3 < \infty$, and for all $\tau,\tau'\in\mathcal{T}$ and $x,x'\in\mathcal{X}$ and $|\mathbf{k}| = p+1$.

Assumption A4. The kernel function $K(\cdot)$ is a product kernel of $k(\cdot)$, which is a symmetric density function with compact support $\mathcal{A}\equiv[-c_k, c_k]$. $\sup_{a\in\mathcal{A}}|k(a)| \leq \overline{c}_1 < \infty$, and $|k(a)-k(a')| \leq \overline{c}_2|a-a'|$ for any $a,a'\in\mathbb{R}$ and some $\overline{c}_2 < \infty$. The functions $H_{\mathbf{j}}(x) = x^{\mathbf{j}}K(x)$ for all $\mathbf{j}$ with $0\leq|\mathbf{j}|\leq 2p+1$ are Lipschitz continuous.

Assumption A5. (i) $G(\cdot)$ is monotone, $\lim_{u\to-\infty}G(u) = 0$, and $\lim_{u\to\infty}G(u) = 1$. (ii) $G(\cdot)$ is three times continuously differentiable with derivatives denoted by $G^{(s)}(\cdot)$ for $s = 1,2,3$. $G(\cdot)$ and its first derivative $G^{(1)}(\cdot)$ are uniformly bounded, and the integrals $\int_{-\infty}^{\infty}\left|G^{(s)}(u)\right|du$, $s = 1,2,3$, are finite. (iii) $g(\cdot)\equiv G^{(1)}(\cdot)$ is symmetric over its support. There exists an integer $q\geq 2$ such that $\int u^s g(u)\,du = 0$ for $s = 1,\cdots,q-1$, and $\int|u^q g(u)|\,du$ is finite. (iv) For some $c_G < \infty$ and $A_G < \infty$, either $G^{(3)}(u) = 0$ for $|u| > A_G$ and for $u,u'\in\mathbb{R}$, $\left|G^{(3)}(u) - G^{(3)}(u')\right| \leq c_G|u-u'|$, or $G^{(3)}(u)$ is differentiable with $|G^{(4)}(u)| \leq c_G$ and for some $\iota_0 > 1$, $|G^{(4)}(u)| \leq c_G|u|^{-\iota_0}$ for all $|u| > A_G$.

Assumption A6. As $n\to\infty$, $h\to 0$, $nh^{3d}/(\log n)^2\to\infty$, $nh^{2(p+1)}\to 0$, $\lambda_n\to 0$, $n\lambda_n^{2q}\to 0$, $n^2\lambda_n^3 h^{7d/2}/\log n\to\infty$, and $n^3\lambda_n^6 h^{4d}/(\log n)^4\to\infty$.

Assumption A7. (i) $\delta(\tau,x,s)$ is uniformly bounded in $(\tau,x,s)\in\mathcal{T}\times\mathcal{X}\times[0,1]$. $\delta(\tau,x,s)$ is continuously differentiable in τ with uniformly bounded derivatives on $\mathcal{T}\times\mathcal{X}\times[0,1]$. (ii) Let $m_{0t\tau}\equiv m_0(\tau,X_t)$, and $\overline{f}_{\lceil n\pi\rfloor}(x)\equiv n^{-1}\sum_{s=1}^{\lceil n\pi\rfloor}f_s(x)$. $\frac{1}{n}\sum_{t=1}^{\lceil n\pi\rfloor}f_t(m_{0t\tau}|X_t)\,\delta(\tau,X_t,t/n) - \frac{c_0}{n}\sum_{t=1}^{n}\overline{f}_{\lceil n\pi\rfloor}(X_t)$ $\overline{f}_n^{-1}(X_t)\,f_t(m_{0t\tau}|X_t)\,\delta(\tau,X_t,t/n)\xrightarrow{P}\Delta(\pi,\tau)+o(1)$ uniformly in $(\pi,\tau)\in[0,1]\times\mathcal{T}$.

Assumption A8. (i) There exists a p.d.f. $f(\cdot)$ such that $|f_{nt}(X_t) - f(X_t)|\to 0$ a.s. as $n\to\infty$. (ii) Let $\varepsilon_{t\tau}\equiv Y_t - m_t(\tau,X_t)$ and $F_{n,ts}(\tau,\tau')\equiv E[\mathbf{1}(\varepsilon_{t\tau}\leq 0)\,\mathbf{1}(\varepsilon_{s\tau'}\leq 0)]$. $\lim_{n\to\infty}F_{n,ts}(\tau,\tau') = F_{t-s}(\tau,\tau')$ for all t,s.

Many parts of Assumptions A1 - A4 are similar to those of Masry (1996) and they are typical assumptions to ensure uniform results in nonparametric literature. Some of these assumptions can be relaxed with modifications on the proof. Assumption A1 restricts that the process $\{(Y_t,X_t)\}$ to be strong mixing with mixing rates decaying sufficiently fast. Assumption A2 imposes smoothness conditions on the marginal and conditional density functions, where neither null nor local alternative condition is imposed. The boundedness of the joint p.d.f. facilitates the determination of moments of certain U-statistics. Assumption A3 is required for the establishment of the uniform local Bahadur representation for our local polynomial estimates. Assumption A4 specifies typical conditions on the kernel used in local polynomial quantile regressions. Assumption A5(i) is required because we use $G(\cdot)$ to approximate the indicator function. Assumptions A5(ii)-(iv) specify the smoothness conditions on the function $G(\cdot)$. In particular, Assumption A5(iii) requires that $g(\cdot)$ behaves like a qth order kernel and Assumption A5(iv) is used in studying the remainder term of a third order Taylor expansion. Assumption A6 imposes conditions on the bandwidth. In particular, the condition $nh^{2(p+1)} \to 0$ implies that undersmoothing is required for our tests. Note that the last requirement in A6 implies that $n^{-1/2}h^{-d/2}\sqrt{\log n} = o(\lambda_n)$, i.e., $n\lambda_n^2 h^d / \log n \to \infty$. If we set $h = n^{-1/\gamma_1}$ and $\lambda_n \propto n^{-1/\gamma_2}$, then we need

$$\max\left(\frac{6\gamma_1}{4\gamma_1 - 7d}, \frac{6\gamma_1}{3\gamma_1 - 4d}\right) < \gamma_2 < 2q.$$

When the dimension d of the conditioning variable X_t is small, $q = 2$ will suffice. For example, if $d = 1$, $p = 1$, $q = 2$, $h \propto n^{-1/3.5}$, then one can choose $\gamma_2 \in (42/13, 4)$; if $d = 2$, $p = 3$, $q = 2$, $h \propto n^{-1/7}$, then one can choose $\gamma_2 \in (42/13, 4)$. Assumption A7 gives some properties of the local alternative; it is not minimal but simplifies our proofs. If the triangular array process $\{(Y_t, X_t)\} \equiv \{(Y_{nt}, X_{nt})\}$ satisfies Assumption A8 and H_{1n}, we say it is *asymptotically stationary.*

To proceed, it is worthwhile to specify the notation on the conditional c.d.f. and p.d.f. under the null hypothesis and local alternatives. First, given the triangular array nature of the process $\{(Y_t, X_t)\} \equiv \{(Y_{nt}, X_{nt})\}$, the conditional p.d.f. $f_t(\cdot|X_t)$ and c.d.f. $F_t(\cdot|X_t)$ in Assumption A2 usually depend on both n and t, that is, $f_t(\cdot|X_t) \equiv f_{nt}(\cdot|X_t)$ and $F_t(\cdot|X_t) \equiv F_{nt}(\cdot|X_t)$. An exception occurs when H_0 holds and $f_t(\cdot|X_t)$ and $F_t(\cdot|X_t)$ do not depend on t. In this case, we will write $f_t(\cdot|X_t)$ simply as $f_0(\cdot|X_t)$, which is the conditional p.d.f. associated with the conditional c.d.f. $F_0(\cdot|X_t)$ and the conditional quantile function $m_0(\tau, X_t)$ under the null hypothesis. Second, letting $m_{t\tau} \equiv m(\tau, X_t)$, it is easy to verify that under H_{1n} (see (2.4)) and Assumption A7 (i), we have

$$f_t(m_{t\tau}|X_t) = f_0(m_{t\tau}|X_t) - n^{-1/2} f_0(m_{t\tau}|X_t)^2 \delta_\tau(\tau, x, t/n) + o_P\left(n^{-1/2}\right), \qquad (3.1)$$

where $\delta_\tau(\tau, \cdot, \cdot) \equiv \partial\delta(\tau, \cdot, \cdot)/\partial\tau$. We will use this relationship in the subsequent asymptotic analysis.

3.2 Asymptotic distributions

We first give a general asymptotic result of $S_n(\cdot, \cdot)$ without Assumption A8. The general result helps us better understand the limiting behavior of the process. Let $\varsigma_{nt}(\pi) \equiv \overline{f}_{\lceil n\pi \rfloor}(X_t)\overline{f}_n^{-1}(X_t)$, and define

$$\begin{aligned}
\Gamma_{11}(\pi_1,\pi_2;\tau_1,\tau_2) &\equiv \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{\lceil n\pi_1\rfloor}\sum_{j=1}^{\lceil n\pi_2\rfloor} E\left[\psi_{\tau_1}(\varepsilon_{i\tau_1})\psi_{\tau_2}(\varepsilon_{j\tau_2})\right],\\
\Gamma_{12}(\pi_1,\pi_2;\tau_1,\tau_2) &\equiv \lim_{n\to\infty}\frac{c_0}{n}\sum_{i=1}^{\lceil n\pi_1\rfloor}\sum_{j=1}^{n} E[\psi_{\tau_1}(\varepsilon_{i\tau_1})\varsigma_{nj}(\pi_2)\psi_{\tau_2}(\varepsilon_{j\tau_2})],\\
\Gamma_{21}(\pi_1,\pi_2;\tau_1,\tau_2) &\equiv \lim_{n\to\infty}\frac{c_0}{n}\sum_{i=1}^{n}\sum_{j=1}^{\lceil n\pi_2\rfloor} E[\varsigma_{ni}(\pi_1)\psi_{\tau_1}(\varepsilon_{i\tau_1})\psi_{\tau_2}(\varepsilon_{j\tau_2})],\\
\Gamma_{22}(\pi_1,\pi_2;\tau_1,\tau_2) &\equiv \lim_{n\to\infty}\frac{c_0^2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E[\varsigma_{ni}(\pi_1)\psi_{\tau_1}(\varepsilon_{i\tau_1})\varsigma_{nj}(\pi_2)\psi_{\tau_2}(\varepsilon_{j\tau_2})].
\end{aligned}$$

In the appendix, we demonstrate that the above limits are well defined by using the Davydov inequality (e.g., Hall and Heyde 1980, pp. 277-278). The following theorem establishes the limit distribution of $S_n(\cdot,\cdot)$ under H_0 and H_{1n}.

Theorem 3.1 *(i) Under H_0 and Assumptions A1-A6, $S_n(\cdot,\cdot) \Rightarrow S_\infty(\cdot,\cdot)$ as $n\to\infty$, where $S_\infty(\cdot,\cdot)$ is a zero-mean Gaussian process with covariance kernel*

$$E\left[S_\infty(\pi_1,\tau_1)S_\infty(\pi_2,\tau_2)\right] = \sum_{i=1}^{2}\sum_{j=1}^{2}(-1)^{i+j}\Gamma_{ij}(\pi_1,\pi_2;\tau_1,\tau_2).$$

(ii) Under H_{1n} and Assumptions A1-A7, $S_n(\cdot,\cdot) \Rightarrow S_\infty(\cdot,\cdot)+\Delta(\cdot,\cdot)$ as $n\to\infty$, where $\Delta(\cdot,\cdot)$ is defined in Assumption A7.

The following remarks concern the interpretation and application of Theorem 3.1.

Remark 1. Theorem 3.1(i) indicates that the process $\{S_n(\cdot,\cdot), n\geq 1\}$ converges to a zero-mean Gaussian process under the null hypothesis of no structural change in the conditional distribution. The covariance kernel of the limiting process $\{S_\infty(\cdot,\cdot)\}$ depends on both the dependence structure in the data and the contribution of parameter estimation error as reflected by Γ_{12}, Γ_{21} and Γ_{22}. Consequently $S_\infty(\cdot,\cdot)$ is generally not a Kiefer process. Theorem 3.1(ii) implies that a test based on a continuous functional of the process $\{S_n(\cdot,\cdot), n\geq 1\}$ potentially has non-trivial power in detecting $n^{-1/2}$-local alternatives provided $\Delta(\pi,\tau) \neq 0$ for (π,τ) in a set of positive Lebesgue measure on $[0,1]\times\mathcal{T}$.

Remark 2. In principle one can construct the KS and CM test statistics based upon the process $\{S_n(\cdot,\cdot), n\geq 1\}$. The limiting distributions of these test statistics are not nuisance parameter free and so critical values cannot be tabulated. One may consider proposing a bootstrap procedure that takes into account the joint presence of data dependence structure and parameter estimation error, and imposes the null restriction at the same time. This turns out to be extremely difficult (if possible at all) in our framework. For example, if one follows Corradi and Swanson (2006) and proposes a block bootstrap procedure by resampling from the observed data $\{(X_t,Y_t)\}_{t=1}^n$, the null restriction will not be imposed. A side problem with this type of resampling scheme is that it requires re-estimating the conditional quantile function under the null restriction for each resample of the data, and thus is computationally demanding. On the other hand, if one tries to apply the block or stationary bootstrap procedure to resample $\widehat{u}_{t\tau}^*$ from the quantile residuals $\{\widehat{u}_{t\tau}\}$, then there is no simple way to construct the bootstrapped data for $\{Y_t\}_{t=1}^n$ because $\widehat{m}(\tau, X_t) + \widehat{u}_{t\tau}^*$ depends on τ and cannot be assigned to an object like Y_t^*.

If the triangular array process $\{(Y_t, X_t)\} \equiv \{(Y_{nt}, X_{nt})\}$ satisfies Assumption A8, the result in Theorem 3.1 can be greatly simplified. We summarize the limiting behavior of the process in the following corollary.

Corollary 3.2 *(i) Under H_0 and Assumptions A1-A6 and A8, $S_n(\cdot,\cdot) \Rightarrow S_\infty(\cdot,\cdot)$ as $n \to \infty$, where $S_\infty(\cdot,\cdot)$ is a zero-mean Gaussian process with covariance kernel*

$$E[S_\infty(\pi_1,\tau_1) S_\infty(\pi_2,\tau_2)] = \left[(\pi_1 \wedge \pi_2) - 2c_0\pi_1\pi_2 + c_0^2\pi_1\pi_2\right]\Gamma^0(\tau_1,\tau_2),$$

and $\Gamma^0(\tau_1,\tau_2) \equiv \sum_{s=-\infty}^{\infty}[F_s(\tau_1,\tau_2) - \tau_1\tau_2]$.

(ii) Under H_{1n} and Assumptions A1-A8, $S_n(\cdot,\cdot) \Rightarrow S_\infty(\cdot,\cdot) + \Delta(\cdot,\cdot)$ as $n \to \infty$, where $\Delta(\pi,\tau)$ $=\text{plim}_{n\to\infty}\left[n^{-1}\sum_{t=1}^{\lceil n\pi\rfloor}\delta_{nt}(\tau) - n^{-1}\pi c_0\sum_{t=1}^{n}\delta_{nt}(\tau)\right]$, and $\delta_{nt}(\tau) \equiv f_0(m_{0t\tau}|X_t)\,\delta(\tau, X_t, t/n)$.

Remark 3. Corollary 3.2(i) indicates that even under asymptotic stationarity, the null limit distribution of the process $\{S_n(\cdot,\cdot), n \geq 1\}$ is still not asymptotically pivotal: it depends on both the dependence structure in the data and the contribution of parameter estimation error as reflected by Γ^0 and c_0, respectively. In the proof of Corollary 3.2, we show that under H_0,

$$S_n(\pi,\tau) = \frac{1}{n^{1/2}}\sum_{i=1}^{\lceil n\pi\rfloor}\psi_\tau(\varepsilon_{i\tau}) - \frac{\pi c_0}{n^{1/2}}\sum_{i=1}^{n}\psi_\tau(\varepsilon_{i\tau}) + o_P(1), \tag{3.2}$$

where $o_P(1)$ holds uniformly in $(\pi,\tau) \in [0,1] \times \mathcal{T}$. The second term on the right hand side of (3.2) is of the same probability order as the first term and reflects the effect of parameter estimation error. As typical block bootstrap requires resampling of blocks whose length $l = l(n)$ is of order $o(n)$, this will cause the effect of parameter estimation error to vanish asymptotically in the bootstrap world. As a consequence, this renders the limiting distribution of the bootstrap analog of $S_n(\pi,\tau)$ unable to approximate the null limit distribution of $S_n(\pi,\tau)$ itself - see Theorem 3.5 and more discussions in Section 3.3 for additional studies on this issue. For this reason, we consider the re-centered process $\{S_n^c(\cdot,\cdot), n \geq 1\}$.

The following theorem summarizes the limiting distributions of $\{S_n^c(\cdot,\cdot), n \geq 1\}$ under both the null and a sequence of local alternatives.

Theorem 3.3 *(i) Under H_0 and Assumptions A1-A6 and A8, $S_n^c(\cdot,\cdot) \Rightarrow S_\infty^c(\cdot,\cdot)$ as $n \to \infty$, where $S_\infty^c(\cdot,\cdot)$ is a zero-mean Gaussian process with covariance kernel*

$$E[S_\infty^c(\pi_1,\tau_1) S_\infty^c(\pi_2,\tau_2)] = (\pi_1 \wedge \pi_2 - \pi_1\pi_2)\,\Gamma^0(\tau_1,\tau_2),$$

and $\Gamma^0(\tau_1,\tau_2) = \sum_{s=-\infty}^{\infty}[F_s(\tau_1,\tau_2) - \tau_1\tau_2]$.

(ii) Under H_{1n} and Assumptions A1-A8, $S_n^c(\cdot,\cdot) \Rightarrow S_\infty^c(\cdot,\cdot) + \Delta^c(\cdot,\cdot)$ as $n \to \infty$, where $\Delta^c(\pi,\tau) = \Delta^0(\pi,\tau) - \pi\Delta^0(1,\tau)$ with $\Delta^0(\pi,\tau) = \int_0^\pi \int f_0(m_0(\tau,x)|x) f(x)\,\delta(\tau,x,s)\,dxds$.

Remark 4. Theorem 3.3(i) shows that under the null, the process $\{S_n^c(\cdot,\cdot), n \geq 1\}$ converges to a two-parameter Kiefer process $\{S_\infty^c(\cdot,\cdot)\}$ that is tied-down in the first argument (see Csörgő and Horváth, 1997, p. 320 or p. 384). By the continuous mapping theorem, Theorem 3.3 implies that

$$KS_n \xrightarrow{D} \sup_{\pi\in[0,1]} \sup_{\tau\in\mathcal{T}} |S_\infty^c(\pi,\tau)|, \;\; CM_n \xrightarrow{D} \int_{\mathcal{T}} \int_0^1 S_\infty^c(\pi,\tau)^2 \, d\pi w(d\tau) \;\text{ under } H_0, \text{ and}$$

$$KS_n \xrightarrow{D} \sup_{\pi\in[0,1]} \sup_{\tau\in\mathcal{T}} |S_\infty^c(\pi,\tau) + \Delta^c(\pi,\tau)|, \;\; CM_n \xrightarrow{D} \int_{\mathcal{T}} \int_0^1 [S_\infty^c(\pi,\tau) + \Delta^c(\pi,\tau)]^2 \, d\pi w(d\tau) \;\text{ under } H_{1n}.$$

Thus the tests KS_n and CM_n generally have non-trivial power in detecting Pitman local alternatives that decay to zero at the parametric $n^{-1/2}$-rate. If $\delta(\tau, X_t, t/n)$ is orthogonal to the conditional p.d.f. $f_0(m_0(\tau, X_t)|X_t)$ so that $\lim_{n\to\infty} E[f_0(m_0(\tau, X_t)|X_t)\delta(\tau, X_t, t/n)] = 0$ for essentially all (t,τ), then the tests have no power in detecting such deviations from the null. Similar phenomena occur in both parametric and nonparametric/semiparametric tests for structural changes in the conditional mean regression framework. In the parametric case, if the structural shifts in the finite dimensional parameters are orthogonal to the mean regressor then the residual-based CUSUM test is not consistent (e.g., Ploberger and Krämer, 1992, 1996). In the latter case, Su and Xiao (2008b) and Su and White (2009a) demonstrate their CUSUM-type tests for nonparametric and semiparametric structural changes also lose power in certain directions.

An important class of conditioning models is the case where X_t includes lagged dependent variables. In this case, a valid regression model usually requires that the error process be an m.d.s. For such models, the asymptotic distributions of the tests KS_n and CM_n are free of nuisance parameters under the null hypothesis, as can be seen from the following corollary.

Corollary 3.4 *If $\{\psi_\tau(\varepsilon_{t\tau}), \mathcal{F}_t\}$ forms an m.d.s. for each τ, then the result of Theorem 3.3(i) holds with the simplified covariance kernel*

$$E[S_\infty^c(\pi_1,\tau_1) S_\infty^c(\pi_2,\tau_2)] = (\pi_1 \wedge \pi_2 - \pi_1\pi_2)(\tau_1 \wedge \tau_2 - \tau_1\tau_2).$$

Thus the null limit distributions of the tests KS_n and CM_n are free of nuisance parameter.

In the next subsection, we propose a simulation-based method that provides valid inference of our tests for general models with correlated errors. The simulation method is in the spirit of block bootstrap (e.g., Künsch (1989), Bühlmann (1994)), and can mimic the null limit distributions of our test statistics.

3.3 A simulation-based method

From the results of Theorem 3.3, we see that re-centering removes the effect of preliminary estimation, but not the effect of serial correlation. If the error process of the quantile regression model is not an m.d.s., the asymptotics of the tests KS_n and CM_n are generally not asymptotically pivotal. So the critical values for these tests cannot be tabulated. In this subsection, we propose a simulation method to obtain the simulated p-values for the case with correlated errors.

The proposed simulation method is similar to Inoue (2001) (also see Bühlmann 1994), and has some similarity in the spirit of block bootstrap but differs from the latter in several aspects. In short, we generate a weighted sum of blocks of the generalized residuals. Given a block length $l \equiv l(n)$, we consider blocks with length l of the generalized residuals $\{\tau - G_{\lambda_n}(\widehat{u}_{i\tau})\}$. Let $\{z_j\}_{j=1}^{n-l+1}$ be a sequence of random weights whose properties are specified in Assumption A9 below, we define the following simulated process

$$S_n^*(\pi, \tau) = n^{-1/2} \sum_{j=1}^{\lceil n\pi \rfloor - l+1} z_j \sum_{i=j}^{j+l-1} \{\tau - G_{\lambda_n}(\widehat{u}_{i\tau})\}.$$

Let $S_n^{c*}(\pi, \tau) = S_n^*(\pi, \tau) - \pi S_n^*(1, \tau)$, we construct the bootstrap versions KS_n^* and CM_n^* of KS_n and CM_n based on $S_n^{c*}(\pi, \tau)$. Our purpose is to use the distribution of the bootstrapped process $S_n^{c*}(\pi, \tau)$ to approximate that of $S_n^c(\pi, \tau)$. The requirements on l and z_j's are stated in the next assumption.

Assumption A9. (i) $\{z_j\}_{j=1}^{n-l+1}$ are i.i.d. and independent of the process $\{(Y_t, X_t)\}$. (ii) $E(z_j) = 0$, $E(z_j^2) = 1/l$, and $E(z_j^4) = O(1/l^2)$. (iii) As $n \to \infty$, $l \to \infty$, $l/n^{1/2} \to 0$, and $nh^d/(l \log n) \to \infty$.

The asymptotic property of the bootstrapped process is summarized in the following theorem.

Theorem 3.5 *Suppose Assumptions A1-A9 hold. Then under either H_0 or H_{1n},*
(i) $S_n^*(\cdot,\cdot) \overset{p}{\Rightarrow} S_\infty^0(\cdot,\cdot)$,
(ii) $S_n^{c*}(\cdot,\cdot) \overset{p}{\Rightarrow} S_\infty^c(\cdot,\cdot)$,
where $S_\infty^0(\cdot,\cdot)$ is a zero-mean Gaussian process with covariance kernel $E[S_\infty^0(\pi_1,\tau_1) S_\infty^0(\pi_2,\tau_2)] = (\pi_1 \wedge \pi_2)\Gamma^0(\tau_1,\tau_2)$, and $\Gamma^0(\cdot,\cdot)$ and $S_\infty^c(\cdot,\cdot)$ are defined in Theorem 3.3.

Remark 5. Theorem 3.5 explains why we construct the testing statistic based on the re-centered process $S_n^c(\cdot,\cdot)$ instead of $S_n(\cdot,\cdot)$. Theorem 3.5(i) shows that the limit of the simulated process $\{S_n^*(\cdot,\cdot), n \geq 1\}$ is different from that of the original process $\{S_n(\cdot,\cdot), n \geq 1\}$ under H_0. Intuitively speaking, the $n^{-1/2}$-rate of local alternatives do not affect the limiting distribution of the simulated process, which causes the difference between the two limiting processes under H_0. This occurs because, due to the additional randomness of $\{z_j\}$ and the assumption $l = o(n^{1/2})$, the simulated process is less sensitive than the original process to the presence of parameter estimation error or any perturbation from the null restriction. The difference between $S_\infty^0(\cdot,\cdot)$ and $S_\infty(\cdot,\cdot)$ indicates that one cannot use $\{S_n^*(\cdot,\cdot), n \geq 1\}$ to obtain the simulated p-values.

Remark 6. Theorem 3.5(ii) shows that each re-centered simulated process $\{S_n^{c*}(\cdot,\cdot), n \geq 1\}$ converges weakly to the same null limit process of $\{S_n^c(\cdot,\cdot), n \geq 1\}$, thus providing a valid asymptotic basis for approximating the null limit distributions of test statistics based on $\{S_n^c(\cdot,\cdot)\}$. In practice, we repeat the bootstrap procedure B times to obtain the sequences $\{KS_{n,j}^*\}_{j=1}^B$ and $\{CM_{n,j}^*\}_{j=1}^B$. We reject the null when, for example, $p^* = B^{-1}\sum_{j=1}^B 1(KS_n \leq KS_{n,j}^*)$ is smaller than the desired significance level. Analogously, one can obtain the simulated p values for the CM_n test.

4 Monte Carlo Simulations

In this section we present a small set of Monte Carlo experiments designed to evaluate the finite sample performance of our tests. We first focus on their finite sample performance under the null and then examine their power properties. Finally, we compare our nonparametric quantile regression-based tests with Qu's (2008) parametric quantile regression-based tests.

4.1 Finite sample level

We consider three data generating processes (DGPs) for the level study:

DGP s1. $Y_t = X_t - 0.5X_t^2 + v_t$, $v_t = \sqrt{0.1 + 0.5X_t^2}\zeta_{1t}$,

DGP s2. $Y_t = X_t - 0.5X_t^2 + v_t$, $v_t = \sqrt{\vartheta_t}\zeta_{1t}$, $\vartheta_t = 0.05 + 0.95\ \vartheta_{t-1} + 0.025v_{t-1}^2$,

DGP s3. $Y_t = X_t - 0.5X_t^2 + v_t$, $v_t = \sqrt{\vartheta_t}\zeta_{1t}$, $\vartheta_t = 0.05 + 0.95\ \vartheta_{t-1} + 0.025v_{t-1}^2$,

where $X_t = 0.5 + 0.8X_{t-1} + \zeta_{2t}$ in DGPs s1-s2, $X_t = 0.5 + 0.4X_{t-1} + 0.4X_{t-1}\mathbf{1}\left(t \geq \lceil n/2 \rfloor\right) + \zeta_{2t}$ in DGPs s3, ζ_{1t} are i.i.d. $N(0,1)$, ζ_{2t} are i.i.d. sum of 48 independent random variables each uniformly distributed on $[-0.25, 0.25]$, and the two processes $\{\zeta_{1t}\}$ and $\{\zeta_{2t}\}$ are independent.

Clearly, DGP s1-s3 specify the same conditional mean model. But they are different in other aspects. First, DGP s1 specifies a traditional error process with conditional heteroskedasticity whereas DGP s2-s3 specifies a GARCH$(1,1)$ error process. Secondly, the conditioning variable X_t exhibits a distributional change in DGP s3 but not in s1 and s2. To see the last difference, recall $\mathcal{F}_{t-1} \equiv \sigma\left(X_{t-i}, i \geq 0, Y_{t-j}, j \geq 1\right)$, and $\varepsilon_{t\tau} \equiv Y_t - m_t\left(\tau, X_t\right)$. It is easy to verify that $\{\psi_\tau\left(\varepsilon_{t\tau}\right),\ \mathcal{F}_t\}$ forms an m.d.s. in DGP s1 but not in DGPs s2-s3.

It is worth mentioning that our tests are based on the local polynomial quantile estimates, which typically require compact support of the conditioning variables. That is why the way we generate X_t seems awkward. On the other hand, according to the central limit theorem we can treat ζ_{2t} as being nearly standard normal random variables but with compact support [-12, 12].

To construct the test statistics, we choose the normalized Epanechnikov kernel (with variance 1),

$$K(u) = \frac{3}{4}\left(1 - \frac{1}{5}u^2\right)\mathbf{1}\left(|u| \leq \sqrt{5}\right). \tag{4.1}$$

Since there is no data-driven procedure to choose the bandwidth for quantile regression, to estimate the τth conditional quantile of Y_t given X_t, we may choose a preliminary bandwidth according to the rule of thumb recommended by Yu and Jones (1998):

$$h_{0\tau} = s_X n^{-1/5}\left\{\tau(1-\tau)[\phi\left(\Phi^{-1}(\tau)\right)]^{-2}\right\}^{1/5},$$

where s_X is the standard deviation of X_t, ϕ and Φ are the standard normal p.d.f. and c.d.f., respectively. Since undersmoothing is required for our test, we modify the above choice of bandwidth to

$$h_{0\tau} = s_X n^{-1/\gamma}\left\{\tau(1-\tau)[\phi\left(\Phi^{-1}(\tau)\right)]^{-2}\right\}^{1/5},$$

where $3 < \gamma < 4$. We may study the behavior of our tests with different choices of λ in order to examine the sensitivity of our test to the bandwidth sequence. Robinson (1991, p.448) and Lee

Table 1: Finite sample rejection frequency under the null

Sample size n	DGP	Test	Block size: $l = \lceil cn^{1/4} \rfloor$ $\lambda_n = 0$ c=0.5	$c = 1$	$c = 2$	$\lambda_n = 0.001$ c=0.5	$c = 1$	$c = 2$	$\lambda_n = 0.01$ c=0.5	$c = 1$	$c = 2$
100	s1	KS_n	0.053	0.036	0.018	0.050	0.027	0.009	0.040	0.025	0.011
		CM_n	0.048	0.038	0.026	0.038	0.032	0.018	0.036	0.025	0.018
	s2	KS_n	0.036	0.027	0.017	0.032	0.021	0.009	0.030	0.016	0.007
		CM_n	0.039	0.038	0.025	0.032	0.018	0.010	0.029	0.019	0.008
	s3	KS_n	0.037	0.029	0.019	0.024	0.010	0.007	0.022	0.012	0.008
		CM_n	0.040	0.034	0.024	0.026	0.017	0.019	0.028	0.021	0.015
200	s1	KS_n	0.049	0.037	0.024	0.048	0.034	0.022	0.045	0.034	0.022
		CM_n	0.049	0.036	0.040	0.035	0.034	0.025	0.036	0.034	0.025
	s2	KS_n	0.053	0.046	0.031	0.048	0.035	0.018	0.048	0.034	0.018
		CM_n	0.053	0.047	0.034	0.041	0.037	0.022	0.039	0.036	0.020
	s3	KS_n	0.038	0.036	0.026	0.036	0.029	0.017	0.040	0.036	0.018
		CM_n	0.039	0.037	0.032	0.033	0.033	0.021	0.035	0.032	0.017
400	s1	KS_n	0.040	0.044	0.032	0.046	0.028	0.024	0.044	0.036	0.032
		CM_n	0.042	0.044	0.036	0.038	0.044	0.038	0.040	0.042	0.038
	s2	KS_n	0.070	0.066	0.050	0.064	0.048	0.036	0.064	0.048	0.046
		CM_n	0.064	0.064	0.054	0.058	0.056	0.044	0.056	0.056	0.052
	s3	KS_n	0.074	0.066	0.052	0.054	0.054	0.034	0.062	0.050	0.040
		CM_n	0.056	0.066	0.060	0.042	0.042	0.030	0.050	0.040	0.044

(2003, p.16) propose very similar devices. Note that these choices for $h_{0\tau}$ and the kernel function meet the requirements for our test. Through a preliminary study, we find our bootstrap-based test is not sensitive to the choice of γ when we restrict $\gamma \in (3,4)$. So we fix $\gamma = 3.5$ below when we report the simulation results. We choose $G(\cdot)$ to be the standard normal c.d.f. For the block bootstrap, we generate $\{z_j\}$ independently from $N(0, 1/l)$.

Table 1 reports the empirical rejection frequencies of our tests at the 5% nominal level. We use 1000 replications for sample sizes $n = 100, 200$, and 500 replications for sample size $n = 400$. To obtain the simulated p-values, we use 199 simulation paths for each replication. To see how our tests are sensitive to the choice of block size l and the smoothing parameter λ_n, we set $\lambda_n = 0$, 0.001, and 0.01, and choose $l = \lceil cn^{1/4} \rfloor$ for three choices of c: 0.5, 1, 2. When $\lambda_n = 0$, we effectively replace the approximating function $G_{\lambda_n}(-\widehat{u}_{t\tau})$ by the indicator function $\mathbf{1}(\widehat{u}_{t\tau} \leq 0)$. Table 1 shows that: (a) our tests are robust to different choices of smoothing parameter values λ_n but is a little bit sensitive to the choice of block size l (or equivalently c in the table); (b) when the sample size is small (n =100, 200) our tests tend to be undersized for large values of block sizes; (c) as the sample size increases, the empirical level approaches the nominal level quickly; (d) the behavior of the CM_n test is quite similar to the KS_n test, but the former is slightly less sensitive to the choice of block size.

4.2 Finite sample power

To consider the finite sample power performance of the tests, we consider the following three alternatives:

DGP p1. $Y_t = g_t(X_t) + v_t$, $v_t = \sqrt{0.1 + 0.5X_t^2}\,(1 + \mathbf{1}(t \geq \lceil n\pi_0 \rfloor))\,\zeta_{1t}$,

DGP p2. $Y_t = g_t(X_t) + v_t$, $v_t = \sqrt{\vartheta_t}\zeta_{1t}$, $\vartheta_t = 0.05 + (0.95\ -0.4\delta_2\ \mathbf{1}(t < \lceil n\pi_0 \rfloor))\vartheta_{t-1}\ +0.025v_{t-1}^2$,

DGP p3. $Y_t = g_t(X_t) + v_t$, $v_t = \sqrt{\vartheta_t}\zeta_{1t}$, $\vartheta_t = 0.05 + (0.95\ -0.4\delta_2\ \mathbf{1}(t < \lceil n\pi_0 \rfloor))\vartheta_{t-1}\ +0.025v_{t-1}^2$,

where $g_t(X_t) = X_t - 0.5X_t^2 + \delta_1\ \mathbf{1}(t \geq \lceil n\pi_0 \rfloor)$, and X_t and ζ_{1t} in DGPs p1-p3 are generated as in DGPs s1-s3, respectively.

We consider different values of (δ_1, δ_2) to evaluate the finite sample performance of our tests under the alternatives. Obviously, when $\delta_1 = \delta_2 = 0$, DGPs p1-p3 reduce to DGPs s1-s3. As the values of δ_1 and δ_2 move away from zero, we have increasing magnitude of structural break. We consider two different break ratios, $\pi_0 = 0.25, 0.5$, to examine whether the tests are sensitive to the timing of the break. Also, we consider six different break sizes, (δ_1, δ_2)= $(1,0)$, $(2,0)$, $(0,1)$, $(0,2)$, $(1,1)$, $(2,2)$ to see how the test is sensitive to the size of the breaks. Note that when $\delta_2 = 0$ and δ_1 is nonzero, we have structural change in the location only. Similarly, when $\delta_1 = 0$ and δ_2 is nonzero, we have structural change in the scale only.

Tables 2-4 report the finite sample performance of our tests under the alternatives. To save computing time, here we use 500 replications for each case. Some of the main findings from Tables 2-4 are: (a) The power of the KS_n and CM_n tests are sensitive to the choice of block size l but not that of λ_n. The large value of l tends to decrease the power of the tests. (b) As the break size δ_1 or δ_2 increases, the powers of both tests increase. But for DGP p1, the breaks in the scale may have adverse effect on the detection of the breaks in the location. (c) Other things being equal, the CM_n test tends to be a little bit more powerful than the KS_n test. (d) As expected, it is easiest to detect a break when it occurs at the halfway point, $\pi_0 = 0.5$. (e) For DGP p1, it is much easier for our tests to detect breaks in the location than the scale of the distribution. But this is not the case for DGPs p2 and p3.

4.3 A Comparison with Linear Quantile Regression-Based Tests

We compare our test with Qu's (2008) linear quantile regression-based test where a linear conditional quantile model is specified. To be specific, we focus on the following linear DGP

$$Y_t = \beta_{0t} + \beta_{1t}X_t + (1 + \beta_{2t}X_t)\,v_t, \tag{4.2}$$

where the process $\{v_t, t \geq 1\}$ is independent of the process $\{X_t, t \geq 1\}$. So the τth conditional quantile function of Y_t given X_t is linear in X_t :

$$m_t(\tau, X_t) = \beta_{0t}(\tau) + \beta_{1t}(\tau)\,X_t, \tag{4.3}$$

where $\beta_{0t}(\tau) = \beta_{0t} + F_{v_t}^{-1}(\tau)$, $\beta_{1t}(\tau) = \beta_{1t} + \beta_{2t}F_{v_t}^{-1}(\tau)$, and $F_{v_t}^{-1}(\cdot)$ is the inverse of the distribution function of v_t. Let $\boldsymbol{\beta}_t(\tau) = (\beta_{0t}(\tau), \beta_{1t}(\tau))'$. Qu was interested in testing the null hypothesis

$$H_0^* : \boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}_0(\tau) \text{ for all } t \geq 1 \text{ and for all } \tau \in \mathcal{T}. \tag{4.4}$$

Table 2: Finite sample power at 0.05 nominal level (DGP p1: n=200)

Tests	Break point	Break size	Block size: $l=\lceil cn^{1/4}\rceil$								
			λ_n= 0			λ_n= 0.001			λ_n= 0.01		
	π_0	(δ_1,δ_2)	c=0.5	$c=1$	$c=2$	c=0.5	$c=1$	$c=2$	c=0.5	$c=1$	$c=2$
KS_n	0.25	(1,0)	0.722	0.640	0.490	0.720	0.666	0.500	0.722	0.656	0.492
		(2,0)	0.988	0.984	0.914	0.988	0.976	0.918	0.988	0.978	0.916
		(0,1)	0.170	0.144	0.116	0.152	0.144	0.102	0.158	0.136	0.106
		(0,2)	0.488	0.432	0.332	0.502	0.482	0.326	0.514	0.462	0.336
		(1,1)	0.634	0.578	0.454	0.652	0.578	0.452	0.632	0.572	0.462
		(2,2)	0.924	0.896	0.808	0.924	0.914	0.816	0.926	0.916	0.816
	0.50	(1,0)	0.930	0.894	0.832	0.918	0.892	0.828	0.926	0.900	0.816
		(2,0)	1	1	1	1	1	1	1	1	1
		(0,1)	0.422	0.418	0.338	0.430	0.414	0.321	0.450	0.396	0.314
		(0,2)	0.894	0.886	0.844	0.914	0.898	0.832	0.918	0.894	0.830
		(1,1)	0.924	0.908	0.882	0.928	0.910	0.878	0.930	0.916	0.878
		(2,2)	0.996	1	0.994	1	0.998	0.992	0.998	1	0.990
CM_n	0.25	(1,0)	0.748	0.702	0.626	0.764	0.694	0.614	0.754	0.722	0.622
		(2,0)	0.988	0.982	0.948	0.990	0.984	0.966	0.990	0.986	0.960
		(0,1)	0.252	0.228	0.190	0.230	0.212	0.174	0.238	0.216	0.168
		(0,2)	0.676	0.634	0.580	0.694	0.664	0.532	0.684	0.668	0.530
		(1,1)	0.662	0.612	0.570	0.670	0.622	0.560	0.652	0.620	0.538
		(2,2)	0.930	0.902	0.820	0.926	0.914	0.824	0.922	0.918	0.844
	0.50	(1,0)	0.926	0.906	0.882	0.932	0.928	0.886	0.934	0.912	0.888
		(2,0)	1	1	1	1	1	1	1	1	1
		(0,1)	0.484	0.494	0.480	0.478	0.498	0.446	0.506	0.494	0.452
		(0,2)	0.946	0.944	0.938	0.950	0.946	0.942	0.940	0.950	0.932
		(1,1)	0.916	0.910	0.904	0.916	0.916	0.898	0.916	0.918	0.896
		(2,2)	1	0.998	1	0.998	1	1	1	1	0.996

Table 3: Finite sample power at 0.05 nominal level (DGP p2: n=200)

Tests	Break point π_0	Break size (δ_1,δ_2)	Block size: $l=\lceil cn^{1/4}\rfloor$ $\lambda_n=0$ c=0.5	$c=1$	$c=2$	$\lambda_n=0.001$ c=0.5	$c=1$	$c=2$	$\lambda_n=0.01$ c=0.5	$c=1$	$c=2$
KS_n	0.25	(1,0)	0.752	0.690	0.512	0.756	0.684	0.484	0.762	0.694	0.468
		(2,0)	0.998	0.998	0.990	0.998	0.996	0.990	0.998	0.998	0.988
		(0,1)	0.818	0.684	0.336	0.820	0.650	0.266	0.816	0.652	0.290
		(0,2)	0.938	0.834	0.412	0.946	0.814	0.320	0.946	0.810	0.340
		(1,1)	0.996	0.996	0.988	0.996	0.996	0.984	0.996	0.994	0.978
		(2,2)	0.998	0.998	0.996	0.998	0.998	0.994	0.998	0.998	0.998
	0.50	(1,0)	0.942	0.916	0.866	0.954	0.934	0.850	0.952	0.942	0.834
		(2,0)	1	1	1	1	1	1	1	1	1
		(0,1)	0.904	0.794	0.500	0.922	0.754	0.340	0.920	0.772	0.352
		(0,2)	0.988	0.914	0.500	0.992	0.892	0.422	0.990	0.914	0.426
		(1,1)	1	0.998	0.996	1	0.998	0.996	0.998	0.998	0.998
		(2,2)	1	1	1	1	1	1	1	1	1
CM_n	0.25	(1,0)	0.810	0.774	0.720	0.810	0.786	0.698	0.806	0.782	0.694
		(2,0)	1	1	0.998	1	1	0.992	1	0.998	0.996
		(0,1)	0.938	0.884	0.640	0.944	0.842	0.585	0.928	0.842	0.568
		(0,2)	0.986	0.952	0.806	0.984	0.950	0.712	0.988	0.954	0.704
		(1,1)	0.996	0.996	0.986	0.996	0.996	0.984	0.996	0.996	0.984
		(2,2)	0.998	0.998	0.996	0.998	0.998	0.996	0.998	0.998	0.996
	0.50	(1,0)	0.958	0.950	0.932	0.978	0.958	0.926	0.974	0.958	0.926
		(2,0)	1	1	1	1	1	1	1	1	1
		(0,1)	0.972	0.926	0.736	0.964	0.896	0.642	0.974	0.896	0.638
		(0,2)	0.994	0.982	0.866	0.996	0.972	0.796	0.996	0.976	0.774
		(1,1)	1	0.998	0.998	1	0.998	0.994	1	1	0.998
		(2,2)	1	1	1	1	1	1	1	1	1

Table 4: Finite sample power at 0.05 nominal level (DGP p3: n=200)

Tests	Break point π_0	Break size (δ_1,δ_2)	Block size: $l = \lceil cn^{1/4} \rfloor$ $\lambda_n = 0$ c=0.5	$\lambda_n = 0$ $c = 1$	$\lambda_n = 0$ $c = 2$	$\lambda_n = 0.001$ c=0.5	$\lambda_n = 0.001$ $c = 1$	$\lambda_n = 0.001$ $c = 2$	$\lambda_n = 0.01$ c=0.5	$\lambda_n = 0.01$ $c = 1$	$\lambda_n = 0.01$ $c = 2$
KS_n	0.25	(1,0)	0.680	0.610	0.436	0.736	0.676	0.510	0.746	0.678	0.508
		(2,0)	0.998	0.998	0.988	0.998	0.998	0.994	0.998	0.998	0.992
		(0,1)	0.820	0.730	0.458	0.818	0.712	0.390	0.828	0.710	0.396
		(0,2)	0.938	0.882	0.560	0.948	0.862	0.480	0.952	0.868	0.492
		(1,1)	0.998	0.998	0.990	0.998	0.994	0.998	0.998	0.998	0.998
		(2,2)	1	1	1	1	1	0.998	1	1	0.998
	0.50	(1,0)	0.856	0.812	0.722	0.878	0.852	0.766	0.874	0.844	0.770
		(2,0)	1	1	0.998	1	1	0.998	1	1	1
		(0,1)	0.890	0.814	0.556	0.896	0.778	0.480	0.898	0.802	0.468
		(0,2)	0.976	0.912	0.666	0.968	0.904	0.568	0.964	0.904	0.562
		(1,1)	0.998	0.996	0.988	0.998	0.996	0.984	0.998	0.998	0.982
		(2,2)	1	1	0.998	1	0.998	0.998	1	1	0.998
CM_n	0.25	(1,0)	0.714	0.682	0.592	0.788	0.748	0.658	0.772	0.740	0.646
		(2,0)	1	1	0.988	1	0.998	0.992	1	1	0.994
		(0,1)	0.948	0.918	0.806	0.942	0.890	0.712	0.930	0.900	0.710
		(0,2)	0.992	0.970	0.890	0.990	0.970	0.810	0.994	0.954	0.824
		(1,1)	0.998	0.998	0.988	0.998	0.998	0.976	0.998	0.998	0.988
		(2,2)	1	1	0.998	1	0.998	0.998	1	1	0.998
	0.50	(1,0)	0.896	0.882	0.832	0.930	0.912	0.880	0.938	0.916	0.888
		(2,0)	1	1	1	1	1	1	1	1	1
		(0,1)	0.964	0.900	0.794	0.952	0.878	0.714	0.942	0.884	0.726
		(0,2)	0.996	0.988	0.894	0.992	0.978	0.840	0.992	0.978	0.838
		(1,1)	0.994	0.992	0.988	0.994	0.994	0.982	0.992	0.990	0.982
		(2,2)	1	1	0.998	1	1	0.998	1	1	0.998

Qu (2008) considered two tests for structural changes across quantiles. The first test is based on the subgradient:

$$H_{n\pi}(\widehat{\boldsymbol{\beta}}(\tau)) = (X'X)^{-1/2} \sum_{t=1}^{\lceil n\pi \rfloor} \overline{X}_t \psi_\tau (Y_t - \overline{X}_t' \widehat{\boldsymbol{\beta}}(\tau)),$$

where $\overline{X}_t = (1, X_t')'$, $X = (\overline{X}_1, \cdots \overline{X}_n)'$, and $\widehat{\boldsymbol{\beta}}(\tau)$ is the linear quantile regression estimate of $\boldsymbol{\beta}_0(\tau)$ under H_0^* by using the full sample. The test statistic is defined as

$$DQ_n = \sup_{\tau \in \mathcal{T}} \sup_{\pi \in [0,1]} \left\| H_{n\pi}(\widehat{\boldsymbol{\beta}}(\tau)) - \pi H_{n1}(\widehat{\boldsymbol{\beta}}(\tau)) \right\|_\infty$$

where $\|\cdot\|_\infty$ is the sup norm, i.e., for a generic vector $a = (a_1, \cdots, a_k)$, $\|a\|_\infty = \max(|a_1|, \cdots, |a_k|)$.

Qu's second test statistic is of Wald type. Let $\widehat{\boldsymbol{\beta}}_1(\pi, \tau)$ denote the quantile regression estimate of $\boldsymbol{\beta}_0(\tau)$ using observations up to $\lceil n\pi \rfloor$ for $\pi \in (0,1)$. Let $\widehat{\boldsymbol{\beta}}_2(\pi, \tau)$ denote the quantile regression estimate of $\boldsymbol{\beta}_0(\tau)$ using the remaining observations. Then the Wald test for no structural change across quantiles is given by

$$DW_n = \sup_{\tau \in \mathcal{T}} \sup_{\pi \in \Pi_\epsilon} n \Delta \widehat{\boldsymbol{\beta}}(\pi, \tau)' \widehat{V}(\pi, \tau)^{-1} \Delta \widehat{\boldsymbol{\beta}}(\pi, \tau)$$

where $\Pi_\epsilon = [\epsilon, 1-\epsilon]$ for some small $\epsilon \in (0, 1/2)$, $\Delta \widehat{\boldsymbol{\beta}}(\pi, \tau) = \widehat{\boldsymbol{\beta}}_2(\pi, \tau) - \widehat{\boldsymbol{\beta}}_1(\pi, \tau)$, and $\widehat{V}(\pi, \tau)$ is a consistent estimate of the limiting variance of $\sqrt{n} \Delta \widehat{\boldsymbol{\beta}}(\pi, \tau)$ under H_0^*, i.e.,

$$\operatorname*{plim}_{p \to \infty} \widehat{V}(\pi, \tau) = \tau(1-\tau) \left\{ \frac{1}{\pi} + \frac{1}{1-\pi} \right\} \Omega_{0\tau}, \;\; \Omega_{0\tau} = H_{0\tau}^{-1} J_0 H_{0\tau}^{-1},$$

where $H_{0\tau} = \text{plim}_{n\to\infty} n^{-1} \sum_{t=1}^n f_{Y_t}\left(F_{Y_t}^{-1}(\tau)\right) \overline{X}_t \overline{X}_t'$, $J_0 = \text{plim}_{n\to\infty} n^{-1} \sum_{t=1}^n \overline{X}_t \overline{X}_t'$, and $f_{Y_t}\left(F_{Y_t}^{-1}(\tau)\right)$ is the conditional density function of Y_t evaluated at the τth conditional quantile. The difficult part in implementing Qu's DW_n test is to estimate $f_{Y_t}\left(F_{Y_t}^{-1}(\tau)\right)$. We follow Qu's advice and estimate it by the difference quotient

$$\Delta_{nt} = \frac{2b_n}{\overline{X}_t' \widehat{\boldsymbol{\beta}}(\tau + b_n) - \overline{X}_t' \widehat{\boldsymbol{\beta}}(\tau - b_n)}. \tag{4.5}$$

where $b_n = n^{-1/5} \left\{ 4.5 \phi^4 \left(\Phi^{-1}(\tau) \right) / [2 \Phi^{-1}(\tau)^2 + 1] \right\}^{1/5}$ as recommended by Bofinger (1975). To avoid division by 0 [which may occur when n is small and τ is close to 0 or 1], we replace the denominator in (4.5) by 1e-6 when it is 0.

The asymptotic null distributions of the DQ_n and DW_n test statistics are asymptotically pivotal and Qu (2008) tabulated their critical values for conventional 1%, 5% and 10% tests. Since Qu's DW_n test needs to split the sample into two parts and one cannot estimate the coefficients well in the quantile regression when the sample size is too small, we follow his advice and choose $\Pi_\epsilon = \mathcal{T} = [0.15, 0.85]$ to implement his parametric tests. The number of replications is 1000.

We first evaluate the performance of Qu's test in the case of functional form misspecification. We consider the scenario when the data are generated from nonlinear quantile processes in DGPs s1-s3, but test for the presence of structural change in the conditional distributions using Qu's linear quantile regression test. Table 5 reports the finite sample "level" of these tests. From Table 5, we see that under functional misspecification, the size of the Qu's test tends to be highly distorted and the

Table 5: Finite sample level of Qu's test for DGPs s1-s3

n	DGP\ nominal level	DQ_n			DW_n		
		0.01	0.05	0.10	0.01	0.05	0.10
100	s1	0.015	0.069	0.141	0.322	0.443	0.528
	s2	0.603	0.799	0.880	0.999	1	1
	s3	0.724	0.864	0.916	0.999	1	1
200	s1	0.017	0.081	0.148	0.138	0.252	0.333
	s2	0.761	0.917	0.958	1	1	1
	s3	0.933	0.976	0.987	1	1	1
400	s1	0.017	0.082	0.156	0.064	0.146	0.239
	s2	0.828	0.941	0.968	1	1	1
	s3	0.988	0.998	1	1	1	1

distortion tends to increase as n increases. Exception occurs when the data are generated via DGP s1. For DGP s1, the empirical level of the DQ_n test is also distorted, but is much less severe than the other cases. This may be due to fact that the m.d.s. condition on $\{\psi_\tau(\varepsilon_{t\tau}), \mathcal{F}_t\}$ is satisfied in this case. On the other hand, the empirical level of the DW_n tests improve as the sample size doubles or quadruples. We conjecture this is due to the fact that the DW_n test demands sample splitting and it cannot be well behaved with as small sample sizes as $n = 100$ or 200. Similar observations are found even if the underlying conditional quantile function is linear.

Due to the level distortion of Qu's test in the case of functional misspecification, it is inappropriate to compare the power performance of his test to that of our test. In addition, it is difficult, if possible at all, to calculate the level-adjusted empirical power.

Nevertheless, if we stick to linear conditional quantile functions, we can compare the power performance of the two sets of tests. For simplicity, we consider the following DGP:

$$Y_t = 1 + \{1 + \delta_1 \mathbf{1}(t \geq \lceil n/2 \rfloor)\} X_t + \{1 + [1 + \delta_2 \mathbf{1}(t \geq \lceil n/2 \rfloor)] X_t\} v_t, \tag{4.6}$$

where the v_t's are i.i.d. $t(3)$ (t distribution with 3 degrees of freedom) and X_t are generated as in DGP s1. Clearly, when $\delta_1 = \delta_2 = 0$, there is no structural change in the conditional quantile or distribution function. Any nonzero value of δ_1 indicates a location change in the conditional distribution. Similarly, any nonzero value of δ_2 indicates a scale change in the conditional distribution.

Table 6 compares Qu's test of H_0^* with our test of H_0. To save space, for our nonparametric test, we only report the empirical rejection frequencies for $\lambda_n = 0$. The total number of replications is 1000 for each scenario. When $\delta_1 = \delta_2 = 0$, Table 6 reports the level behavior of both types of tests. Clearly, the levels of both Qu's DQ_n test and our tests behave reasonably well. Like the nonlinear case, the level of Qu's DW_n test is highly distorted for the sample sizes under investigation. When $\delta_1 \neq 0$ or $\delta_2 \neq 0$, Table 6 reports the power behavior of both types of tests. Surprisingly, our nonparametric test performs almost as well as, if not better, than the Qu's DQ_n test except for the case when n is too small and the block size is too large ($n = 100$ and $c = 2$).

Table 6. Finite sample rejection frequencies under linear DGP in (4.2) (nominal level: 0.05)

		Qu's test		Our test: block size $l = \lceil cn^{1/4} \rfloor$.					
					KS_n			CM_n	
n	(δ_1, δ_2)	DQ_n	DW_n	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
100	(0,0)	0.036	0.278	0.062	0.048	0.025	0.057	0.057	0.050
	(1,0)	0.194	0.372	0.324	0.264	0.186	0.366	0.337	0.132
	(2,0)	0.734	0.658	0.832	0.778	0.644	0.872	0.852	0.811
	(0,1)	0.040	0.403	0.087	0.073	0.045	0.087	0.070	0.073
	(0,2)	0.077	0.570	0.173	0.143	0.093	0.135	0.138	0.136
	(1,1)	0.163	0.456	0.298	0.243	0.172	0.297	0.272	0.289
	(2,2)	0.490	0.684	0.674	0.615	0.474	0.663	0.642	0.592
200	(0,0)	0.026	0.100	0.047	0.052	0.037	0.049	0.051	0.044
	(1,0)	0.466	0.317	0.613	0.585	0.552	0.647	0.641	0.619
	(2,0)	0.985	0.966	0.991	0.988	0.973	0.999	0.998	0.992
	(0,1)	0.059	0.218	0.143	0.127	0.113	0.106	0.111	0.122
	(0,2)	0.155	0.476	0.380	0.354	0.317	0.333	0.342	0.337
	(1,1)	0.398	0.379	0.594	0.568	0.506	0.579	0.582	0.550
	(2,2)	0.908	0.872	0.968	0.956	0.940	0.968	0.959	0.947

5 Concluding Remarks

In this paper we propose a test for structural change in conditional distributions. It is based upon the local polynomial quantile regression, and thus does not need to specify any conditional mean, variance, or quantile regression model. Moreover, it has non-trivial power to detect deviations from the null at the parametric rate $n^{-1/2}$. To implement our test, one needs to choose the block size to obtain the simulated p-values. It is important to derive a data-driven procedure to select the block size, which requires the study of the trade-off between size and power under a sequence of local alternatives. Another potential extension of our test is to allow fixed breaks in the distribution of the conditioning variable. We leave these for future research.

References

Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821-856.

Andrews, D. W. K., and Ploberger, W.,1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383-1414.

Andrews, D. W. K., and Pollard, D., 1994. An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review* 62, 119-132.

Bai, J. 1994. Weak convergence of the sequential empirical processes of residuals in ARMA models. *Annals of Statistics* 22, 2051-2061.

Bai, J., Lumsdaine, R. L., and Stock, J., 1998. Testing for and dating breaks in integrated and cointegrated time series. *Review of Economic Studies* 64, 395-432.

Bofinger, E., 1975. Estimation of a density function using order statistics. *Australian Journal of Statistics* 17. 1-7.

Brown, R. L., Durbin, J. and Evans, J. M., 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B* 37, 149-163.

Bühlmann, P., 1994. Blockwise bootstrapped empirical process for stationary sequences. *Annals of Statistics* 22, 995-1012.

Chaudhuri, P., 1991. Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics* 19, 760-777.

Corradi, V. and Swanson, N. R., 2006. Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779-806.

Csensov, N. N., 1955. Limit theorems for some classes of random functions. *Conference on Theory of Probability and Mathematical Statistics*, Erevan.

Csörgő, M. and Horváth, L., 1997. *Limit Theorems in Change-Point Analysis.* John Wiley & Sons, New York.

Elliott, G. and Müller, U., 2006. Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies* 73, 907–940.

Fan, J., 1992. Design-adaptive nonparametric regression. *Journal of American Statistical Association* 87, 998-1004.

Fan, J., and Gijbels I., 1996. *Local Polynomial Modeling and Its Applications.* Chapman & Hall, London.

Fan, J., Hu, T-C., and Truong, Y. K., 1994. Robust non-parametric function estimation. *Scandinavian Journal of Statistics* 21, 433-446.

Ghysels, E. and Hall, A., 1990. A test for structural stability of Euler conditions parameters estimated via the generalized method of moments estimator. *International Economic Review* 31, 355-364.

Giné, E. and Zinn, J., 1990. Bootstrapping general empirical measures. *Annals of Probability* 18, 851-869.

Hall, P. and Heyde, C. C., 1980. *Martingale Limit Theory and Its Applications.* Academic Precess, San Diego.

Hansen, B. E., 2000. Testing for structural change in conditional models. *Journal of Econometrics* 97, 93-115.

Inoue, A., 2001. Testing for distributional change in time series. *Econometric Theory 17*, 156-187.

Koenker, R., Ng, P. T. and Portnoy, S., 1994. Quantile smoothing splines. *Biometrika*, 81 673-680.

Künsch, H. R., 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217-1241.

Lee, S., 2003. Efficient semiparametric estimation of partially linear quantile regression model. *Econometric Theory* 19, 1-31.

Lee, S. and Na, S., 2004. A nonparametric test for the change of the density function in strong mixing process. *Statistics & Probability Letters* 66, 25-34.

Li, H., 2008. Estimation and testing of Euler equation models with time-varying reduced-form coefficients, *Journal of Econometrics* 142, 425-448.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis* 17, 571-599.

Nyblom, J., 1989. Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223–230.

Page, E. S., 1955. A test for change in a parameter occurring at an unknown point. *Biometrika* 42, 523-527.

Picard, D., 1985. Testing and estimating change-point in time series. *Advances in Applied Probability* 17, 841-867.

Ploberger, W. and Krämer, W., 1992. The CUSUM test with OLS residuals. *Econometrica* 56, 1355-1369.

Ploberger, W. and Krämer, W., 1996. A trend resistant test for structural change based on ols residuals. *Journal of Econometrics* 70, 175-186.

Pollard, D., 1990. *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 2.

Qu, Z., 2008. Testing for structural change in regression quantiles. *Journal of Econometrics* 146, 170-184.

Robinson, P. M., 1991. Consistent nonparametric entropy-based testing. *Review of Economic Studies* 58, 437-453.

Rozenholc, Y., 2001. Nonparametric tests of change-points with tapered data. *Journal of Time Series Analysis* 22, 13-43.

Sowell, F., 1996. Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica* 64, 1085–1107.

Su, L., and White, H., 2009a. Testing structural change in partially linear models. Forthcoming in *Econometric Theory.*

Su, L., and White, H., 2009b. Local polynomial quantile regression under nonstationary data: uniform Bahadur representation with applications to testing conditional independence. Working paper, School of Economics, Singapore Management Univ.

Su, L. and Xiao, Z., 2008a. Testing for parametric stability in quantile regression models. *Statistics and Probability Letters* 78, 2768-2775.

Su, L. and Xiao, Z., 2008b. Testing structural change in time-series nonparametric regression models. *Statistics and Its Interface* 1, 347-366.

Van der Vaart, W., 1998. *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Van der Vaart, A. and Wellner, J. A., 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer, New York.

White, H., 2001. *Asymptotic Theory for Econometricians.* 2nd edition. Academic Press, San Diego.

Yu, K. and Jones, M. C., 1998. Local linear quantile regression. *Journal of American Statistical Association* **93**, 228-237.

APPENDIX

A Some Useful Propositions

In this appendix, we prove some propositions that are used in the proof of Theorem 3.1 under Assumptions A1-A7. The main results are proved in the next appendix. Recall $\varepsilon_{i\tau} \equiv Y_i - m_i(\tau, X_i)$, $u_{i\tau} \equiv Y_i - m_0(\tau, X_i)$, and $\widehat{u}_{i\tau} \equiv Y_i - \widehat{m}(\tau, X_i)$. Let $\mu_{ix} \equiv \mu\left((X_i - x)/h\right)$, $K_{ix} \equiv K\left((x - X_i)/h\right)$, and $m_{i\tau} \equiv m_i(\tau, X_i)$. Let E_{X_i} and E_i denote expectation conditional on X_i and $\mathcal{F}_{i-1}$, respectively, where recall $\mathcal{F}_{i-1} = \sigma(X_{i-t}, t \geq 0; Y_{i-s}, s > 1)$. We write $A_n \simeq B_n$ to signify that $A_n = B_n[1 + o_P(1)]$ as $n \to \infty$. We use C to signify a generic constant whose exact value may vary from case to case.

Proposition A.1 *(i)* $E_i\left[G\left(-\varepsilon_{i\tau}/\lambda_n\right) - \mathbf{1}(\varepsilon_{i\tau} \leq 0)\right] = O_P\left(\lambda_n^q\right)$ *uniformly in* τ;

(ii) $E_i\left|G\left(-\varepsilon_{i\tau}/\lambda_n\right) - \mathbf{1}\left(\varepsilon_{i\tau} \leq 0\right)\right|^s = O_P\left(\lambda_n\right)$ *uniformly in* τ *for any* $s \geq 1$;

(iii) $E\left|G\left(-\varepsilon_{i\tau}/\lambda_n\right) - \mathbf{1}\left(\varepsilon_{i\tau} \leq 0\right)\right|^s = O\left(\lambda_n\right)$ *for any* $s \geq 1$;

(iv) $\lambda_n^{-1} E_{X_i}\left[G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right] = f_i\left(m_{i\tau}|X_i\right) + O_P\left(\lambda_n^q\right)$;

(v) $\sup_{\tau\in\mathcal{T}} \frac{1}{n\lambda_n} \sum_{i=1}^n \left|G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right| = O_P(1)$;

(vi) $\sup_{\tau\in\mathcal{T}} \frac{1}{n\lambda_n^2} \sum_{i=1}^n \left|G^{(2)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right| = O_P(1 + n^{-1/2}\lambda_n^{-3/2}\sqrt{\log n})$;

(vii) $\sup_{\tau\in\mathcal{T}} \frac{1}{n\lambda_n^3} \sum_{i=1}^n \left|G^{(3)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right| = O_P(1 + n^{-1/2}\lambda_n^{-5/2}\sqrt{\log n})$.

Proof. The proof is similar to that of Propositions B.1 and B.3 of Su and White (2009b). ■

Proposition A.2 $\sup_{\pi\in[0,1]} \sup_{\tau\in\mathcal{T}} \left|n^{-1/2} \sum_{i=1}^{\lceil n\pi\rfloor} \left\{\mathbf{1}\left(\varepsilon_{i\tau} \leq 0\right) - G\left(-\varepsilon_{i\tau}/\lambda_n\right)\right\}\right| = o_P(1)$.

Proof. Decompose $n^{-1/2} \sum_{i=1}^{\lceil n\pi\rfloor} \left\{\mathbf{1}\left(\varepsilon_{i\tau} \leq 0\right) - G\left(-\varepsilon_{i\tau}/\lambda_n\right)\right\} = \mathbb{V}_{1n}(\pi,\tau) + \mathbb{V}_{2n}(\pi,\tau)$, where

$$\begin{aligned}
\mathbb{V}_{1n}(\pi,\tau) &\equiv n^{-1/2} \sum_{i=1}^{\lceil n\pi\rfloor} \left\{\mathbf{1}\left(\varepsilon_{i\tau} < 0\right) - G\left(-\varepsilon_{i\tau}/\lambda_n\right) - E_i[\mathbf{1}\left(\varepsilon_{i\tau} < 0\right) - G\left(-\varepsilon_{i\tau}/\lambda_n\right)]\right\}, \\
\mathbb{V}_{2n}(\pi,\tau) &\equiv n^{-1/2} \sum_{i=1}^{\lceil n\pi\rfloor} E_i[\mathbf{1}\left(\varepsilon_{i\tau} < 0\right) - G\left(-\varepsilon_{i\tau}/\lambda_n\right)].
\end{aligned}$$

By Proposition A.1(i) and Assumption A6, $\sup_{\pi\in[0,1]} \sup_{\tau\in\mathcal{T}} \left|\mathbb{V}_{2n}(\pi,\tau)\right| = O_P\left(n^{1/2}\lambda_n^q\right) = o_P(1)$. To study $\mathbb{V}_{1n}(\pi,\tau)$, we partition the compact set $\mathcal{T}$ by $\overline{n}_1 - 1$ points $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{\overline{n}_1} = \overline{\tau}$ such that $|\tau_j - \tau_{j-1}| = n^{-(1/2+\epsilon_1)}$ for $\epsilon_1 > 0$. Let $\mathcal{T}_j \equiv (\tau_{j-1}, \tau_j]$ for $j = 1, \cdots, \overline{n}_1$. Let $\tau \in \mathcal{T}_j$. By the monotonicity of $\mathbf{1}\left(\varepsilon_{i\tau} < 0\right)$ and $G\left(-\varepsilon_{i\tau}/\lambda_n\right)$ in τ, we have

$$\begin{aligned}
\mathbb{V}_{1n}(\pi,\tau) \leq\ & n^{-1/2} \sum_{i=1}^{\lceil n\pi\rfloor} \left\{\mathbf{1}(\varepsilon_{i\tau_j} \leq 0) - G(-\varepsilon_{i\tau_{j-1}}/\lambda_n) - E_i[\mathbf{1}(\varepsilon_{i\tau_j} \leq 0) - G(-\varepsilon_{i\tau_{j-1}}/\lambda_n)]\right\} \\
& + n^{-1/2} \sum_{i=1}^{\lceil n\pi\rfloor} \left\{E_i[G\left(-\varepsilon_{i\tau}/\lambda_n\right) - G(-\varepsilon_{i\tau_{j-1}}/\lambda_n) - \mathbf{1}(\varepsilon_{i\tau} \leq 0) + \mathbf{1}(\varepsilon_{i\tau_j} \leq 0)]\right\},
\end{aligned}$$

and

$$
\begin{aligned}
\mathbb{V}_{1n}(\pi,\tau) &\geq n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}\left\{\mathbf{1}(\varepsilon_{i\tau_{j-1}}\leq 0)-G(-\varepsilon_{i\tau_j}/\lambda_n)-E_i[\mathbf{1}(\varepsilon_{i\tau_{j-1}}\leq 0)-G(-\varepsilon_{i\tau_j}/\lambda_n)]\right\} \\
&\quad +n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}\left\{E_i[G(-\varepsilon_{i\tau}/\lambda_n)-G(-\varepsilon_{i\tau_j}/\lambda_n)-\mathbf{1}(\varepsilon_{i\tau}\leq 0)+\mathbf{1}(\varepsilon_{i\tau_{j-1}}\leq 0)]\right\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\sup_{0\leq\pi\leq 1}\sup_{\tau\in\mathcal{T}}|\mathbb{V}_{1n}(\pi,\tau)| \\
\leq\ & \max_{1\leq r\leq n}\max_{1\leq j\leq\overline{n}_1}\left|n^{-1/2}\sum_{i=1}^{r}\left\{\mathbf{1}(\varepsilon_{i\tau_j}\leq 0)-G(-\varepsilon_{i\tau_{j-1}}/\lambda_n)-E_i[\mathbf{1}(\varepsilon_{i\tau_j}\leq 0)-G(-\varepsilon_{i\tau_{j-1}}/\lambda_n)]\right\}\right| \\
&+\max_{1\leq r\leq n}\max_{1\leq j\leq\overline{n}_1}\left|n^{-1/2}\sum_{i=1}^{r}\left\{\mathbf{1}(\varepsilon_{i\tau_{j-1}}\leq 0)-G(-\varepsilon_{i\tau_j}/\lambda_n)-E_i[\mathbf{1}(\varepsilon_{i\tau_{j-1}}\leq 0)-G(-\varepsilon_{i\tau_j}/\lambda_n)]\right\}\right| \\
&+\max_{1\leq r\leq n}\max_{1\leq j\leq\overline{n}_1}\sup_{\tau\in\mathcal{T}_j}\left|n^{-1/2}\sum_{i=1}^{r}\left\{E_i[G(-\varepsilon_{i\tau}/\lambda_n)-G(-\varepsilon_{i\tau_{j-1}}/\lambda_n)-\mathbf{1}(\varepsilon_{i\tau}\leq 0)+\mathbf{1}(\varepsilon_{i\tau_j}\leq 0)]\right\}\right| \\
&+\max_{1\leq r\leq n}\max_{1\leq j\leq\overline{n}_1}\sup_{\tau\in\mathcal{T}_j}\left|n^{-1/2}\sum_{i=1}^{r}\left\{E_i[G(-\varepsilon_{i\tau}/\lambda_n)-G(-\varepsilon_{i\tau_j}/\lambda_n)-\mathbf{1}(\varepsilon_{i\tau}\leq 0)+\mathbf{1}(\varepsilon_{i\tau_{j-1}}\leq 0)\right\}\right| \\
\equiv\ & V_{n1}+V_{n2}+V_{n3}+V_{n4},\ \text{say.}
\end{aligned}
$$

Let $\xi_{t,j}=\mathbf{1}(\varepsilon_{t\tau_j}\leq 0)-G(-\varepsilon_{t\tau_{j-1}}/\lambda_n)-E_t[\mathbf{1}(\varepsilon_{t\tau_j}\leq 0)-G(-\varepsilon_{t\tau_{j-1}}/\lambda_n)]$. Let $\gamma>1$. Noting that $\left\{\xi_{t,j},\mathcal{F}_t\right\}$ is an m.d.s. by construction, it follows from the Chebyshev and Doob inequalities (e.g., Hall and Heyde, 1980, p. 15) that

$$
\begin{aligned}
P(V_{n1}>\epsilon) &\leq \overline{n}_1\max_{1\leq j\leq\overline{n}_1}P\left(\max_{1\leq r\leq n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{r}\xi_{t,j}\right|>\epsilon\right) \\
&\leq \frac{2\gamma\overline{n}_1\epsilon^{-2\gamma}}{2\gamma-1}\max_{1\leq j\leq\overline{n}_1}E\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_{t,j}\right\|^{2\gamma}.
\end{aligned}
$$

By the Rosenthal inequality (e.g., Hall and Heyde (1980, pp. 23-24)) and Proposition A.1(ii),

$$
\begin{aligned}
E\left\|\sum_{i=1}^{n}\xi_{i,j}\right\|^{2\gamma} &\leq C\left\{E\left[\sum_{t=1}^{n}E\left(\left\|\xi_{t,j}\right\|^2|\mathcal{F}_{t-1}\right)\right]^{\gamma}+\sum_{t=1}^{n}E\left\|\xi_{t,j}\right\|^{2\gamma}\right\} \\
&\leq C\left\{n^{\gamma}\lambda_n^{\gamma}+n\lambda_n\right\}\ \text{uniformly in } j.
\end{aligned}
$$

Thus

$$
\begin{aligned}
P(V_{n1}>\epsilon) &\leq C\epsilon^{-2\gamma}\overline{n}_1\left\{\lambda_n^{\gamma}+n^{1-\gamma}\lambda_n\right\} \\
&= C\epsilon^{-2\gamma}n^{\frac{1}{2}+\epsilon_1}\left\{\lambda_n^{\gamma}+n^{1-\gamma}\lambda_n\right\}=o(1)
\end{aligned}
$$

for sufficiently large γ. That is, $V_{n1}=o_P(1)$. By the same token, we can show that $V_{n2}=o_p(1)$. By Proposition A.1(i), it can be shown that $V_{n3}=O_P(n^{1/2}(\lambda_n^q+n^{-(1/2+\epsilon_1)}))=o_P(1)$. Similarly, $V_{n4}=o_P(1)$. ■

Proposition A.3 *Under* H_{1n}, $n^{-1/2}h^{-d}\sum_{i=1}^{n}\left[\mathbf{1}\left(\varepsilon_{i\tau}\leq 0\right)-\mathbf{1}\left(u_{i\tau}\leq 0\right)\right]\mu_{ix}K_{ix}=n^{-1}h^{-d}\sum_{i=1}^{n}$ $f_i(m_0\left(\tau,X_i\right)|X_i)\delta(\tau,X_i,i/n)\mu_{ix}K_{ix}+o_P\left(1\right)$ *uniformly in* $(\tau,x)\in\mathcal{T}\times\mathcal{X}$.

Proof. Decompose $n^{-1/2}h^{-d}\sum_{i=1}^{n}\left[\mathbf{1}\left(\varepsilon_{i\tau}\leq 0\right)-\mathbf{1}\left(u_{i\tau}\leq 0\right)\right]\mu_{ix}K_{ix}=\mathbb{V}_{3n}\left(\tau,x\right)+\mathbb{V}_{4n}\left(\tau,x\right),$ where

$$\begin{aligned}\mathbb{V}_{3n}\left(\tau,x\right) &\equiv n^{-1/2}h^{-d}\sum_{i=1}^{n}\left\{\mathbf{1}\left(\varepsilon_{i\tau}\leq 0\right)-\mathbf{1}\left(u_{i\tau}\leq 0\right)-\tau+E_{X_i}[\mathbf{1}\left(u_{i\tau}\leq 0\right)]\right\}\mu_{ix}K_{ix},\\ \mathbb{V}_{4n}\left(\tau,x\right) &\equiv n^{-1/2}h^{-d}\sum_{i=1}^{n}\left\{\tau-E_{X_i}[\mathbf{1}\left(u_{i\tau}\leq 0\right)]\right\}\mu_{ix}K_{ix}.\end{aligned}$$

Analogously to the proof of Theorem 2.1 in Su and White (2009b), one can show that $\mathbb{V}_{3n}\left(\tau,x\right)=o_P\left(1\right)$ uniformly in $(\tau,x)\in\mathcal{T}\times\mathcal{X}$. Let $\Delta_n\left(\tau,x\right)\equiv n^{-1}h^{-d}\sum_{i=1}^{n}f_i\left(m_0\left(\tau,X_i\right)|X_i\right)\delta\left(\tau,X_i,i/n\right)\mu_{ix}K_{ix}$. By the Taylor expansion and Assumptions A2(ii), A6, and A7(i),

$$\begin{aligned}&\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}}\left\|\mathbb{V}_{4n}\left(\tau,x\right)-\Delta_n\left(\tau,x\right)\right\|\\ =&\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}}\left\|n^{-1/2}h^{-d}\sum_{i=1}^{n}\left[F_i\left(m_{i\tau}|X_i\right)-F_i\left(m_0\left(\tau,X_i\right)|X_i\right)\right]\mu_{ix}K_{ix}-\Delta_n\left(\tau,x\right)\right\|\\ =&\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}}\left\|n^{-1}h^{-d}\sum_{i=1}^{n}\left[f_i\left(m_{i\tau}^{*}|X_i\right)-f_i\left(m_0\left(\tau,X_i\right)|X_i\right)\right]\delta\left(\tau,X_i,i/n\right)\mu_{ix}K_{ix}\right\|\\ \leq&\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}}n^{-3/2}h^{-d}\sum_{i=1}^{n}C_2\left(X_i\right)\delta^2\left(\tau,X_i,i/n\right)\left\|\mu_{ix}K_{ix}\right\|\\ \leq&\ C\sup_{x\in\mathcal{X}}n^{-3/2}h^{-d}\sum_{i=1}^{n}\left\|\mu_{ix}K_{ix}\right\|=o_P\left(1\right),\end{aligned}$$

where $m_{i\tau}^{*}$ lies between $m_{i\tau}$ and $m_0\left(\tau,X_i\right)$. The result follows. ■

Proposition A.4 *Under* H_{1n}, $n^{-1/2}\lambda_n^{-1}\sum_{i=1}^{\lceil n\pi\rfloor}G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\left(\widehat{u}_{i\tau}-\varepsilon_{i\tau}\right)=-c_0n^{-1/2}\sum_{i=1}^{n}\overline{f}_{\lceil n\pi\rfloor}\left(X_i\right)$ $\overline{f}_n^{-1}\left(X_i\right)\psi_\tau\left(\varepsilon_{i\tau}\right)+\Delta\left(\pi,\tau\right)+o_P\left(1\right)$ *uniformly in* $\left(\pi,\tau\right)$.

Proof. Let $\delta_{ni\tau}\equiv\delta\left(\tau,X_i,i/n\right),\ g_{ni\tau}\equiv\lambda_n^{-1}G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right),\ \mu_{ij}\equiv\mu\left((X_i-X_j)/h\right),$ and $\mathbb{S}_n\left(\pi,\tau\right)\equiv n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}g_{ni\tau}\left(\widehat{u}_{i\tau}-\varepsilon_{i\tau}\right)$. Noting that $u_{i\tau}=\varepsilon_{i\tau}+n^{-1/2}\delta_{ni\tau}$, we have

$$\begin{aligned}\mathbb{S}_n\left(\pi,\tau\right) &= n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}g_{ni\tau}\left(\widehat{u}_{i\tau}-u_{i\tau}\right)+n^{-1}\sum_{i=1}^{\lceil n\pi\rfloor}g_{ni\tau}\delta_{ni\tau}\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lceil n\pi\rfloor}f_i\left(m_{i\tau}|X_i\right)\left(\widehat{u}_{i\tau}-u_{i\tau}\right)+\frac{1}{\sqrt{n}}\sum_{i=1}^{\lceil n\pi\rfloor}\left[g_{ni\tau}-E\left(g_{ni\tau}|X_i\right)\right]\left(\widehat{u}_{i\tau}-u_{i\tau}\right)\\ &\quad+\frac{1}{n}\sum_{i=1}^{\lceil n\pi\rfloor}g_{ni\tau}\delta_{ni\tau}+o_P\left(1\right)\\ &\equiv \mathbb{S}_{n1}\left(\pi,\tau\right)+\mathbb{S}_{n2}\left(\pi,\tau\right)+\mathbb{S}_{n3}\left(\pi,\tau\right)+o_P\left(1\right),\end{aligned}\tag{A.1}$$

where $o_P(1)$ holds uniformly in (π, τ) by Proposition A.1(iv) and Assumption A6. By Proposition 2.1, Assumptions A7(i) and (3.1), we have that under H_{1n},

$$\begin{aligned}
\widehat{u}_{i\tau} - u_{i\tau} &= -e_1' H_n(\tau, X_i)^{-1} \frac{1}{nh^d} \sum_{j=1}^{n} \psi_\tau(u_{j\tau}) \mu_{ji} K_{ij} [1 + o_P(1)] + o_P(n^{-1/2}) \\
&= -f_0(m_{i\tau}|X_i)^{-1} \overline{f}_n(X_i)^{-1} e_1' \mathbb{H}^{-1} \frac{1}{nh^d} \sum_{j=1}^{n} \psi_\tau(u_{j\tau}) \mu_{ji} K_{ij} [1 + o_P(1)] + o_P(n^{-1/2})
\end{aligned}$$

where both $o_P(1)$ and $o_P(n^{-1/2})$ hold uniformly in i and τ and $\overline{f}_n(x) \equiv n^{-1} \sum_{j=1}^n f_j(x)$. In addition, noting that under H_{1n} $\widehat{u}_{i\tau} - u_{i\tau} = O_p\left(n^{-1/2} h^{-d/2} \sqrt{\log n}\right)$ and $f_i(m_{i\tau}|X_i) = f_0(m_{i\tau}|X_i) + O_P(n^{-1/2})$ both uniformly in i and τ, it follows that

$$\mathbb{S}_{n1}(\pi, \tau) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{\lceil n\pi \rfloor} \overline{f}_n(X_i)^{-1} e_1' \mathbb{H}^{-1} \frac{1}{nh^d} \sum_{j=1}^{n} \psi_\tau(u_{j\tau}) \mu_{ji} K_{ij} [1 + o_P(1)] + o_P(n^{-1/2}).$$

Let $\alpha_{1i}(\tau) \equiv e_1' \mathbb{H}^{-1} \frac{1}{nh^d} \sum_{j=1}^n \psi_\tau(\varepsilon_{j\tau}) \mu_{ji} K_{ij}$, and $\alpha_{2i}(\tau) \equiv e_1' \mathbb{H}^{-1} \frac{1}{nh^d} \sum_{j=1}^n [\mathbf{1}(\varepsilon_{j\tau} \leq 0) - \mathbf{1}(u_{j\tau} \leq 0)] \times \mu_{ji} K_{ij}$. Then we have

$$\begin{aligned}
\mathbb{S}_{n1}(\pi, \tau) &\simeq -\frac{1}{\sqrt{n}} \sum_{i=1}^{\lceil n\pi \rfloor} \overline{f}_n(X_i)^{-1} \alpha_{1i}(\tau) - \frac{1}{\sqrt{n}} \sum_{i=1}^{\lceil n\pi \rfloor} \overline{f}_n(X_i)^{-1} \alpha_{2i}(\tau) + o_P(1) \\
&\equiv -\mathbb{S}_{n1}^{(1)}(\pi, \tau) - \mathbb{S}_{n1}^{(2)}(\pi, \tau) + o_P(1), \text{ say.} \qquad \text{(A.2)}
\end{aligned}$$

We first study $\mathbb{S}_{n1}^{(1)}(\pi, \tau)$. Let $w_i \equiv (X_i', Y_i)'$, $\varsigma_{ij} \equiv h^{-d} \overline{f}_n(X_i)^{-1} e_1' \mathbb{H}^{-1} \mu_{ji} K_{ij}$, $\varphi_{1\tau}(w_i, w_j) \equiv [\varsigma_{ij} - E_{X_i}(\varsigma_{ij})] \psi_\tau(\varepsilon_{j\tau})$ and $\varphi_2(w_i, w_j) \equiv E_{X_i}(\varsigma_{ij}) \psi_\tau(\varepsilon_{j\tau})$. Then

$$\begin{aligned}
\mathbb{S}_{n1}^{(1)}(\pi, \tau) &= \frac{1}{n^{3/2}} \sum_{i=1}^{\lceil n\pi \rfloor} \sum_{j=1}^{n} \varphi_{1\tau}(w_i, w_j) + \frac{1}{n^{3/2}} \sum_{i=1}^{\lceil n\pi \rfloor} \sum_{j=1}^{n} \varphi_{2\tau}(w_i, w_j) \\
&\equiv \mathbb{S}_{n11}^{(1)}(\pi, \tau) + \mathbb{S}_{n12}^{(1)}(\pi, \tau). \qquad \text{(A.3)}
\end{aligned}$$

Partition $\mathcal{T}$ as before by $n_1 - 1$ points $\underline{\tau} = \tau_0 < \tau_1 < \cdots < \tau_{n_1} = \overline{\tau}$, but we now require $\tau_s - \tau_{s-1} = h^{d/2} / \log n$. Let $\mathcal{T}_s = [\tau_{s-1}, \tau_s]$ for $s = 1, \cdots, n_1$. Then

$$\begin{aligned}
& \sup_{\pi \in [0,1]} \sup_{\tau \in \mathcal{T}} \left| \mathbb{S}_{n11}^{(1)}(\pi, \tau) \right| \\
\leq & \sup_{\pi \in [0,1]} \max_{1 \leq s \leq n_1} \left| \mathbb{S}_{n11}^{(1)}(\pi, \tau_s) \right| + \sup_{\pi \in [0,1]} \max_{1 \leq s \leq n_1} \sup_{\tau \in \mathcal{T}_s} \left| \mathbb{S}_{n11}^{(1)}(\pi, \tau) - \mathbb{S}_{n11}^{(1)}(\pi, \tau_s) \right|. \qquad \text{(A.4)}
\end{aligned}$$

Write

$$\begin{aligned}
\mathbb{S}_{n11}^{(1)}(\pi, \tau) &= \frac{1}{n^{3/2}} \sum_{i=1}^{\lceil n\pi \rfloor} \varphi_{1\tau}(w_i, w_i) + \frac{1}{n^{3/2}} \sum_{1 \leq i < j \leq \lceil n\pi \rfloor} \varphi_{1\tau}(w_i, w_j) \\
&\quad + \frac{1}{n^{3/2}} \sum_{1 \leq j < i \leq \lceil n\pi \rfloor} \varphi_{1\tau}(w_i, w_j) - \frac{1}{n^{3/2}} \sum_{\lceil n\pi \rfloor + 1 \leq i < j \leq n} \varphi_{1\tau}(w_i, w_j) \\
&\equiv \mathbb{S}_{n11a}^{(1)}(\pi, \tau) + \mathbb{S}_{n11b}^{(1)}(\pi, \tau) + \mathbb{S}_{n11c}^{(1)}(\pi, \tau) - \mathbb{S}_{n11d}^{(1)}(\pi, \tau).
\end{aligned}$$

It can be shown that $\sup_{\pi\in[0,1]}\max_{1\le s\le n_1}|\mathbb{S}^{(1)}_{n11a}(\pi,\tau_s)|=O_P(n^{-1/2}h^{-d})=o_P(1)$ by Assumption A6. Let $\mathbb{U}(r,\tau)=r^{-2}\sum_{1\le i<j\le r}\varphi_{1\tau}(w_i,w_j)$. Then $\mathbb{S}_{n11b}(r/n,\tau)=r^2n^{-3/2}\mathbb{U}(r,\tau)$. For $\varphi_{1\tau}$, define M_{n1s} $(s=1,2,3,4)$ and M_{n2s} $(s=1,2,3)$ as in Lemma C.3 of Su and White (2009b). It is easy to verify that

$$\begin{aligned} M_{n11} &= M_{n12}=O\left(h^{-d\eta}\right),\ M_{n13}=M_{n14}=O(h^{-d(1+\eta)}), \\ M_{n21} &= O(h^{-2d}),\ M_{n22}=O(h^{-2d}),\ \text{and } M_{n23}=O(h^{-3d}), \end{aligned}$$

which implies that $E\left[\mathbb{U}(n,\tau)\right]^4=O\left(n^{-4}(h^{-4d\eta/(4+\eta)}+h^{-2d})\right)$. Given $\epsilon>0$. By Lemma C.3(i) of Su and White (2009b), the Bonferonni and Markov inequalities, and Assumptions A1 and A6,

$$\begin{aligned} & P\left(\sup_{\pi\in[0,1]}\max_{1\le s\le n_1}\left|\mathbb{S}^{(1)}_{n11b}(\pi,\tau_s)\right|\ge\epsilon\right) \\ =\ & P\left(\max_{1\le s\le n_1}\max_{1\le r\le n}\frac{r^2}{n^{3/2}}\left|\mathbb{U}(r,\tau_s)\right|\ge\epsilon\right)\le n_1\max_{1\le s\le n_1}\sum_{r=1}^{n}P\left(\left|\mathbb{U}(r,\tau_s)\right|\ge n^{3/2}\epsilon/r^2\right) \\ \le\ & n_1O\left(n^{-6}(h^{-4d\eta/(4+\eta)}+h^{-2d})\right)\sum_{r=1}^{n}r^4=O\left(n^{-1}(h^{-4d\eta/(4+\eta)}+h^{-2d})h^{-d/2}\log n\right)=o(1). \end{aligned}$$

Thus $\sup_{\pi\in[0,1]}\max_{1\le s\le n_1}\left|\mathbb{S}^{(1)}_{n11b}(\pi,\tau_s)\right|=o_P(1)$. Similarly, $\sup_{\pi\in[0,1]}\max_{1\le s\le n_1}\left|\mathbb{S}^{(1)}_{n11l}(\pi,\tau_s)\right|=o_P(1)$, $l=c,d$. Consequently

$$\sup_{\pi\in[0,1]}\max_{1\le s\le n_1}\left|\mathbb{S}^{(1)}_{n11}(\pi,\tau_s)\right|=o_P(1). \tag{A.5}$$

Next,

$$\left|\mathbb{S}^{(1)}_{n11}(\pi,\tau)-\mathbb{S}^{(1)}_{n11}(\pi,\tau_s)\right|\le\frac{1}{n^{1/2}}\sum_{j=1}^{n}\left|\psi_\tau(\varepsilon_{j\tau})-\psi_\tau(\varepsilon_{j\tau_s})\right|\left|\chi_{\lceil n\pi\rfloor,j}\right|$$

where $\chi_{\lceil n\pi\rfloor,j}\equiv\chi_{\lceil n\pi\rfloor}(X_j)$, and $\chi_{\lceil n\pi\rfloor}(x)\equiv\frac{1}{nh^d}\sum_{i=1}^{\lceil n\pi\rfloor}e_1'\mathbb{H}^{-1}\{\overline{f}_n(X_i)^{-1}\mu((x-X_i)/h)K((x-X_i)/h)$ $-E[\overline{f}_n(X_i)^{-1}\mu((x-X_i)/h)K((x-X_i)/h)]\}$. We can show that $\chi_{\lceil n\pi\rfloor}(x)=O_P(n^{-1/2}h^{-d/2}\sqrt{\log n})$ uniformly in (π,x). Also, it is straightforward to show that $\frac{1}{n^{1/2}}\sum_{j=1}^{n}|\psi_\tau(\varepsilon_{j\tau})-\psi_\tau(\varepsilon_{j\tau_s})|=O_P\left(n^{1/2}h^{d/2}/\log n\right)$ uniformly in τ such that $|\tau-\tau_s|=O\left(h^{d/2}/\log n\right)$. Consequently

$$\sup_{\pi\in[0,1]}\max_{1\le s\le n_1}\sup_{\tau\in\mathcal{T}_s}\left|\mathbb{S}^{(1)}_{n11}(\pi,\tau)-\mathbb{S}^{(1)}_{n11}(\pi,\tau_s)\right|=o_P(1). \tag{A.6}$$

Combining (A.4), (A.5), and (A.6) yields

$$\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}\left|\mathbb{S}^{(1)}_{n11}(\pi,\tau)\right|=o_P(1). \tag{A.7}$$

For the second term in (A.3), we have

$$\begin{aligned} \mathbb{S}^{(1)}_{n12}(\pi,\tau) &= \frac{1}{n^{3/2}h^d}\sum_{i=1}^{\lceil n\pi\rfloor}\sum_{j=1}^{n}E_{X_i}\left[\overline{f}_n(X_i)^{-1}e_1'\mathbb{H}^{-1}\mu_{ji}K_{ij}\right]\psi_\tau(\varepsilon_{j\tau}) \\ &= \frac{1}{n^{3/2}}\sum_{i=1}^{\lceil n\pi\rfloor}\sum_{j=1}^{n}\int f_i(X_j+hv)\,\overline{f}_n(X_j+hv)^{-1}e_1'\mathbb{H}^{-1}\mu(v)K(v)\,dv\ \psi_\tau(\varepsilon_{j\tau}) \\ &\simeq \frac{c_0}{n^{1/2}}\sum_{j=1}^{n}\overline{f}_{\lceil n\pi\rfloor}(X_j)\,\overline{f}_n^{-1}(X_j)\,\psi_\tau(\varepsilon_{j\tau}). \end{aligned} \tag{A.8}$$

Now we study $\mathbb{S}_{n1}^{(2)}(\pi,\tau)$ defined in (A.2). By Proposition A.3

$$\begin{aligned}
\mathbb{S}_{n1}^{(2)}(\pi,\tau) &= \frac{1}{n^2h^d}\sum_{i=1}^{\lceil n\pi\rfloor}\overline{f}_n(X_i)^{-1}e_1'\mathbb{H}^{-1}\sum_{j=1}^{n}f_j(m_0(\tau,X_j)|X_j)\,\delta(\tau,X_j,j/n)\,\mu_{ji}K_{ij}+o_P(1)\\
&= \frac{c_0}{n}\sum_{j=1}^{n}\overline{f}_{\lceil n\pi\rfloor}(X_j)\,\overline{f}_n(X_j)^{-1}f_j(m_0(\tau,X_j)|X_j)\,\delta(\tau,X_j,j/n)+o_P(1). \quad \text{(A.9)}
\end{aligned}$$

Combining (A.2), (A.3), and (A.7)-(A.9), we have that uniformly in (π,τ)

$$\begin{aligned}
\mathbb{S}_{n1}(\pi,\tau) &= -\frac{c_0}{n^{1/2}}\sum_{j=1}^{n}\overline{f}_{\lceil n\pi\rfloor}(X_j)\,\overline{f}_n^{-1}(X_j)\,\psi_\tau(\varepsilon_{j\tau})\\
&\quad -\frac{c_0}{n}\sum_{j=1}^{n}\overline{f}_{\lceil n\pi\rfloor}(X_j)\,\overline{f}_n(X_j)^{-1}f_j(m_0(\tau,X_j)|X_j)\,\delta(\tau,X_j,j/n)+o_P(1). \quad \text{(A.10)}
\end{aligned}$$

Analogously to the proof of $\mathbb{S}_{n11b}^{(1)}(\pi,\tau)$, we can show that

$$\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}|\mathbb{S}_{n2}(\pi,\tau)| = O_P\left(n^{-2}(\lambda_nh^d)^{-6\eta/(\eta+6)}+n^{-2}(\lambda_nh^d)^{-3})h^{-d/2}\log n\right)=o_P(1). \quad \text{(A.11)}$$

Now, write $\mathbb{S}_{n3}(\pi,\tau)=\frac{1}{n}\sum_{i=1}^{\lceil n\pi\rfloor}[g_{ni\tau}-E_{X_i}(g_{ni\tau})]\,\delta_{ni\tau}+\frac{1}{n}\sum_{i=1}^{\lceil n\pi\rfloor}E_{X_i}(g_{ni\tau})\,\delta_{ni\tau}$. It is easy to show that the first term is $o_P(1)$ uniformly in (π,τ). The second term is $\frac{1}{n}\sum_{i=1}^{\lceil n\pi\rfloor}f_i(m_{i\tau}|X_i)\,\delta_{ni\tau}+o_P(1)$ by Proposition A.1(iv). Hence

$$\mathbb{S}_{n3}(\pi,\tau)=\frac{1}{n}\sum_{i=1}^{\lceil n\pi\rfloor}f_i(m_{i\tau}|X_i)\,\delta_{ni\tau}+o_P(1) \text{ uniformly in } (\pi,\tau). \quad \text{(A.12)}$$

Combining (A.1) and (A.10)-(A.12) yields the desired result under Assumption A7. ■

Proposition A.5 *Under H_{1n}, $n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}\{G(-\varepsilon_{i\tau}/\lambda_n)-G(-\widehat{u}_{i\tau}/\lambda_n)\}=-c_0n^{-1/2}\sum_{i=1}^{n}\overline{f}_{\lceil n\pi\rfloor}(X_i)$ $\overline{f}_n^{-1}(X_i)\,\psi_\tau(\varepsilon_{i\tau})$ $+\Delta(\pi,\tau)+o_P(1)$ uniformly in (π,τ).*

Proof. By the Taylor expansion

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{\lceil n\pi\rfloor}\{G(-\varepsilon_{i\tau}/\lambda_n)-G(-\widehat{u}_{i\tau}/\lambda_n)\}\\
&= \frac{1}{\sqrt{n}\lambda_n}\sum_{i=1}^{\lceil n\pi\rfloor}G^{(1)}(-\varepsilon_{i\tau}/\lambda_n)(\widehat{u}_{i\tau}-\varepsilon_{i\tau})+\frac{1}{2\sqrt{n}\lambda_n^2}\sum_{i=1}^{\lceil n\pi\rfloor}G^{(2)}(-\varepsilon_{i\tau}/\lambda_n)(\widehat{u}_{i\tau}-\varepsilon_{i\tau})^2\\
&\quad +\frac{1}{6\sqrt{n}\lambda_n^3}\sum_{i=1}^{\lceil n\pi\rfloor}G^{(3)}(-\varepsilon_{i\tau}/\lambda_n)(\widehat{u}_{i\tau}-\varepsilon_{i\tau})^3+R_n(\pi,\tau)\\
&\equiv T_{n1}(\pi,\tau)+T_{n2}(\pi,\tau)+T_{n3}(\pi,\tau)+R_n(\pi,\tau), \quad \text{(A.13)}
\end{aligned}$$

where $R_n(\pi,\tau)\equiv(1/6)n^{-1/2}\lambda_n^{-3}\sum_{i=1}^{\lceil n\pi\rfloor}\left[G^{(3)}(-\overline{u}_{i\tau}/\lambda_n)-G^{(3)}(-u_{i\tau}/\lambda_n)\right](\widehat{u}_{i\tau}-\varepsilon_{i\tau})^3$ with $\overline{u}_{i\tau}$ lying between $\widehat{u}_{i\tau}$ and $\varepsilon_{i\tau}$. By Proposition A.4, it suffices to show the last three terms in (A.13) are uniformly $o_P(1)$. Standard arguments (as in Masry (1996)) show that $\vartheta_n\equiv\max_{1\le i\le n}\sup_{\tau\in\mathcal{T}}|\widehat{u}_{i\tau}-$

$\varepsilon_{i\tau}| = O_P(n^{-1/2}h^{-d/2}\sqrt{\log n} + n^{-1/2}) = O_P(n^{-1/2}h^{-d/2}\sqrt{\log n})$ under H_{1n}. By Proposition A.1(vi) and Assumption A6,

$$\begin{aligned}\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}|T_{n2}(\pi,\tau)| &\leq \max_{1\leq i\leq n}\sup_{\tau\in\mathcal{T}} n^{1/2}|\widehat{u}_{i\tau}-\varepsilon_{i\tau}|^2 \sup_{\tau\in\mathcal{T}}\left\{\frac{1}{2n\lambda_n^2}\sum_{i=1}^n\left|G^{(2)}(-\varepsilon_{i\tau}/\lambda_n)\right|\right\}\\ &= O_P(n^{-1/2}h^{-d}\log n)O_P(1+n^{-1/2}\lambda_n^{-3/2}\sqrt{\log n}) = o_P(1).\end{aligned}$$

Similarly, by Proposition A.1(vii) and Assumption A6 we have $\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}|T_{n3}(\pi,\tau)| = O_P(n^{-1}h^{-3d/2}(\log n)^{3/2})\ O_P(1+n^{-1/2}\lambda_n^{-5/2}(\log n)^{1/2}) = o_P(1)$. Assumption A5(iv) implies that for all $|\varepsilon-\varepsilon^*|\leq\delta\leq A_g$,

$$|G^{(3)}(\varepsilon^*)-G^{(3)}(\varepsilon)|\leq\delta G^*(\varepsilon).$$

In fact, one chooses $G^*(\varepsilon) = c_G\mathbf{1}(|\varepsilon|\leq 2A_G)$ if $G^{(3)}(\varepsilon)$ has compact support and is Lipschitz continuous, and chooses $G^*(\varepsilon) = c_G\mathbf{1}(|\varepsilon|\leq 2A_G)+|\varepsilon-A_G|^{-\gamma_0}\mathbf{1}(|\varepsilon|>2A_G)$. In each case, $G^*(\varepsilon)$ is bounded and integrable and behaves like the kernel function $K(\cdot)$. Noting that $\vartheta_n/\lambda_n\leq A_G$ with probability approaching one (w.p.a. 1), we have that w.p.a. 1, $\left|G^{(3)}(-\overline{u}_{i\tau}/\lambda_n)-G^{(3)}(-\varepsilon_{i\tau}/\lambda_n)\right|\leq \vartheta_n\lambda_n^{-1}G^*(-\varepsilon_{i\tau}/\lambda_n)$, and $\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}R_n(\pi,\tau)\leq\frac{n^{1/2}\vartheta_n^4\lambda_n^{-3}}{6}\sup_{\tau\in\mathcal{T}}\frac{1}{n\lambda_n}\sum_{i=1}^n G^*(-\varepsilon_{i\tau}/\lambda_n) = O_P\left(n^{-3/2}\lambda_n^{-3}h^{-2d}(\log n)^2\right) = o_P(1)$ because $\sup_{\tau\in\mathcal{T}}\frac{1}{n\lambda_n}\sum_{i=1}^n G^*(-\varepsilon_{i\tau}/\lambda_n) = O_P(1)$ following the proof of Proposition A.1(v). ■

The following proposition is used in the proof of Theorem 3.5.

Proposition A.6 *Under H_{1n}, $l^{-1/2}\sum_{j=i}^{i+l-1}[G(-\varepsilon_{j\tau}/\lambda_n)-G(-\widehat{u}_{j\tau}/\lambda_n)] = o_P(1)$ uniformly in $\tau\in\mathcal{T}$.*

Proof. The proof is analogous to that of Proposition A.5. The major difference is to evaluate $T_{n1i}(\tau)\equiv(\sqrt{l}\lambda_n)^{-1}\sum_{j=i}^{i+l-1}G^{(1)}(-\varepsilon_{j\tau}/\lambda_n)(\widehat{u}_{j\tau}-\varepsilon_{j\tau})$, the analog of $T_{n1}(\pi,\tau)$ in the proof of Proposition A.5. By Proposition A.1(v) and Assumption A9(iii),

$$\begin{aligned}\sup_{\tau\in\mathcal{T}}|T_{n1i}(\tau)| &\leq \sqrt{l}\max_{1\leq j\leq n}\sup_{\tau\in\mathcal{T}}|\widehat{u}_{j\tau}-\varepsilon_{j\tau}|\sup_{\tau\in\mathcal{T}}\frac{1}{l\lambda_n}\sum_{j=i}^{i+l-1}\left|G^{(1)}(-\varepsilon_{j\tau}/\lambda_n)\right|\\ &= \sqrt{l}O_P(n^{-1/2}h^{-d/2}\sqrt{\log n})O_P(1) = o_P(1).\end{aligned}$$

Then the result follows. ■

B Proof of the Main Results in Section 3

Proof of Theorem 3.1

We only prove part (ii) of Theorem 3.1 since the proof of part (i) is much simpler. We decompose

$S_n(\pi,\tau)$ as follows: $S_n(\pi,\tau) = S_{n1}(\pi,\tau) + S_{n2}(\pi,\tau) + S_{n3}(\pi,\tau)$, where

$$\begin{aligned}
S_{n1}(\pi,\tau) &\equiv n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}\{\tau - \mathbf{1}(\varepsilon_{i\tau}<0)\},\\
S_{n2}(\pi,\tau) &\equiv n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}\{\mathbf{1}(\varepsilon_{i\tau}<0) - G(-\varepsilon_{i\tau}/\lambda_n)\},\\
S_{n3}(\pi,\tau) &\equiv n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}\{G(-\varepsilon_{i\tau}/\lambda_n) - G(-\widehat{u}_{i\tau}/\lambda_n)\}.
\end{aligned}$$

By Propositions A.2 and A.5, we have that uniformly in (π,τ)

$$\begin{aligned}
S_{n2}(\pi,\tau) &= o_P(1),\\
S_{n3}(\pi,\tau) &= -\frac{c_0}{n^{1/2}}\sum_{i=1}^{n}\overline{f}_{\lceil n\pi\rfloor}(X_i)\,\overline{f}_n^{-1}(X_i)\,\psi_\tau(\varepsilon_{i\tau}) + \Delta(\pi,\tau) + o_P(1),
\end{aligned}$$

where recall $\overline{f}_{\lceil n\pi\rfloor}(x) \equiv n^{-1}\sum_{i=1}^{\lceil n\pi\rfloor} f_i(x)$. It follows that

$$S_n(\pi,\tau) = \overline{S}_n(\pi,\tau) + \Delta(\pi,\tau) + o_P(1), \tag{B.1}$$

where

$$\overline{S}_n(\pi,\tau) = \frac{1}{n^{1/2}}\sum_{i=1}^{\lceil n\pi\rfloor}\psi_\tau(\varepsilon_{i\tau}) - \frac{c_0}{n^{1/2}}\sum_{i=1}^{n}\overline{f}_{\lceil n\pi\rfloor}(X_i)\,\overline{f}_n^{-1}(X_i)\,\psi_\tau(\varepsilon_{i\tau}). \tag{B.2}$$

It suffices to show that $\overline{S}_n(\cdot,\cdot) \Rightarrow S_\infty(\cdot,\cdot)$, where $S_\infty(\cdot,\cdot)$ is defined in Theorem 3.1.

Step 1. We first show the convergence of the sample covariance kernel to the specified covariance kernel. Write

$$\begin{aligned}
&E\left[\overline{S}_n(\pi_1,\tau_1)\,\overline{S}_n(\pi_2,\tau_2)\right]\\
=\;& \frac{1}{n}\sum_{i=1}^{\lceil n\pi_1\rfloor}\sum_{j=1}^{\lceil n\pi_2\rfloor} E\left[\psi_{\tau_1}(\varepsilon_{i\tau_1})\,\psi_{\tau_2}(\varepsilon_{j\tau_2})\right]\\
&+\frac{c_0^2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E[\overline{f}_{\lceil n\pi_1\rfloor}(X_i)\,\overline{f}_n^{-1}(X_i)\,\psi_{\tau_1}(\varepsilon_{i\tau_1})\,\overline{f}_{\lceil n\pi_2\rfloor}(X_j)\,\overline{f}_n^{-1}(X_j)\,\psi_{\tau_2}(\varepsilon_{j\tau_2})]\\
&-\frac{c_0}{n}\sum_{i=1}^{\lceil n\pi_1\rfloor}\sum_{j=1}^{n} E[\psi_{\tau_1}(\varepsilon_{i\tau_1})\,\overline{f}_{\lceil n\pi_2\rfloor}(X_j)\,\overline{f}_n^{-1}(X_j)\,\psi_{\tau_2}(\varepsilon_{j\tau_2})]\\
&-\frac{c_0}{n}\sum_{i=1}^{n}\sum_{j=1}^{\lceil n\pi_2\rfloor} E[\overline{f}_{\lceil n\pi_1\rfloor}(X_i)\,\overline{f}_n^{-1}(X_i)\,\psi_{\tau_1}(\varepsilon_{i\tau_1})\,\psi_{\tau_2}(\varepsilon_{j\tau_2})]\\
\equiv\;& S_{n11}(\pi_1,\pi_2;\tau_1,\tau_2) + S_{n22}(\pi_1,\pi_2;\tau_1,\tau_2) - S_{n12}(\pi_1,\pi_2;\tau_1,\tau_2) - S_{n21}(\pi_1,\pi_2;\tau_1,\tau_2).
\end{aligned}$$

By the Davydov inequality (e.g., Hall and Heyde, 1980, pp. 277-278),

$$\frac{1}{n}\sum_{i=1}^{\lceil n\pi_1\rfloor}\sum_{j=1}^{\lceil n\pi_2\rfloor}\left|E\left[\psi_{\tau_1}(\varepsilon_{i\tau_1})\,\psi_{\tau_2}(\varepsilon_{j\tau_2})\right]\right| \le \frac{4}{n}\sum_{i=1}^{\lceil n\pi_1\rfloor}\sum_{j=1}^{\lceil n\pi_2\rfloor}\alpha(|i-j|) \le 4\sum_{s=0}^{\infty}\alpha(s) < \infty.$$

It follows that $S_{n11}(\pi_1, \pi_2; \tau_1, \tau_2)$ is absolutely convergent, and

$$S_{n11}(\pi_1, \pi_2; \tau_1, \tau_2) = \frac{1}{n} \sum_{i=1}^{\lceil n\pi_1 \rfloor} \sum_{j=1}^{\lceil n\pi_2 \rfloor} E\left[\psi_{\tau_1}(\varepsilon_{i\tau_1}) \psi_{\tau_2}(\varepsilon_{j\tau_2})\right] \to \Gamma_{11}(\pi_1, \pi_2; \tau_1, \tau_2).$$

Similarly, it can be shown that $S_{nst}(\pi_1, \pi_2; \tau_1, \tau_2) \to \Gamma_{st}(\pi_1, \pi_2; \tau_1, \tau_2)$ for $(s,t) = (1,2), (2,1)$, and $(2,2)$, where Γ_{st} are defined before Theorem 3.1.

Step 2. We now establish the convergence of finite dimensional distribution. Write $\overline{S}_n(\pi, \tau) = n^{-1/2} \sum_{i=1}^{n} [\mathbf{1}(i \leq \lceil n\pi \rfloor) - c_0 \overline{f}_{\lceil n\pi \rfloor}(X_i) \overline{f}_n^{-1}(X_i)] \psi_\tau(\varepsilon_{i\tau})$. Fix $k \geq 1$, $\omega \equiv (\omega_1, \cdots, \omega_k)' \in \mathbb{R}^k$ with $\|\omega\| = 1$, and $(\pi_1, \cdots, \pi_k) \in [0,1]^k$. Let $\varsigma_{n,i} = \sum_{j=1}^{k} \omega_j [\mathbf{1}(i \leq \lceil n\pi_j \rfloor) - c_0 \overline{f}_{\lceil n\pi_j \rfloor}(X_i) \overline{f}_n^{-1}(X_i)] \psi_{\tau_j}(\varepsilon_{i\tau_j})$. By the Cramér-Wold device, it suffices to show that

$$\sum_{j=1}^{k} \omega_j \overline{S}_n(\pi_j, \tau_j) = n^{-1/2} \sum_{i=1}^{n} \varsigma_{n,i}$$

is asymptotically normally distributed. By Assumptions A3 (i)-(ii), the $\varsigma_{n,i}$'s are bounded constants, i.e., $\sup_{n \geq 1} \max_{1 \leq i \leq n} |\varsigma_{n,i}| \leq \overline{c} < \infty$. Clearly, $E[\varsigma_{n,i}] = 0$. By the Davydov inequality, it is easy to show that $s_n^2 \equiv \mathrm{Var}(n^{-1/2} \sum_{i=1}^{n} \varsigma_{n,i}) = \sum_{i=1}^{n} E(\varsigma_{n,i}^2) + 2 \sum_{1 \leq i < j \leq n}^{n} E[\varsigma_{n,i} \varsigma_{n,j}]$ is bounded by a finite constant. Our strong mixing condition is stronger than that required in Theorem 5.20 of White (2001). It follows from the same theorem that $\sum_{j=1}^{k} \omega_j \overline{S}_n(\pi_j, \tau_j) \xrightarrow{d} N(0, \lim_{n \to \infty} s_n^2)$.

Step 3. We show the uniform tightness of the process $\overline{S}_n(\cdot, \cdot)$. Write $\overline{S}_n(\pi, \tau) = n^{-1/2} \sum_{i=1}^{n} s_{ni}(\pi) \psi_\tau(\varepsilon_{i\tau})$, where

$$s_{ni}(\pi) = \mathbf{1}(i \leq \lceil n\pi \rfloor) - c_0 \overline{f}_{\lceil n\pi \rfloor}(X_i) \overline{f}_n^{-1}(X_i).$$

Like Picard (1985) and Rozenholc (2001), we use the following Csensov (1955) tightness criterion for continuous processes. Let $\mathcal{C}([0,1] \times \mathcal{T}; \mathbb{R})$ denote the space of real-valued continuous processes indexed by $(\pi, \tau) \in [0,1] \times \mathcal{T}$. A family of process $\{\mathcal{S}_n(\cdot, \cdot)\}$ of $\mathcal{C}([0,1] \times \mathcal{T}; \mathbb{R})$ is tight if the following four conditions are satisfied: (i) $\{\mathcal{S}_n(0,0)\}$ is tight in $\mathbb{R}$; (ii) $\{\mathcal{S}_n(0, \cdot)\}$ is tight in $\mathcal{C}(\mathcal{T}; \mathbb{R})$; (iii) $\{\mathcal{S}_n(\cdot, \underline{\tau})\}$ is tight in $\mathcal{C}([0,1]; \mathbb{R})$; (iv) there exist constants $C > 0$, $\gamma_1 > 1$, $\gamma_2 > 0$, and a measure ν defined on $\mathcal{T}$ absolutely continuous with respect to the Lebesgue measure and such that for all $B \equiv [\pi_1, \pi_2] \times [\tau_1, \tau_2]$ included in $[0,1] \times \mathcal{T}$, we have

$$E|\mathcal{S}_n(B)|^{\gamma_1} \leq C(\pi_2 - \pi_1) \nu(\tau_2 - \tau_1)^{\gamma_2}$$

where $\mathcal{S}_n(B) \equiv \mathcal{S}_n(\pi_2, \tau_2) - \mathcal{S}_n(\pi_2, \tau_1) - \mathcal{S}_n(\pi_1, \tau_2) + \mathcal{S}_n(\pi_1, \tau_1)$.

Noting that the process $\{\overline{S}_n(\cdot, \cdot)\}$ is not continuous with respect to either of its two coordinates, we cannot apply the above criterion directly. Nevertheless, we can define a process $\{L_n(\cdot, \cdot)\}$ that belongs to $\mathcal{C}([0,1] \times \mathcal{T}; \mathbb{R})$ as follows: $L_n(0, \tau) = 0$, $L_n(k/n, \tau) = n^{-1/2} \sum_{i=1}^{n} s_{ni}(k/n) \{E[G(\varepsilon_{i\tau}/\lambda_n)] - G(\varepsilon_{i\tau}/\lambda_n)\}$, and for $\pi \in [0,1]$

$$L_n(\pi, \tau) = L_n\left(\frac{\lceil n\pi \rfloor}{n}, \tau\right) + (n\pi - \lceil n\pi \rfloor) \left[L_n\left(\frac{\lceil n\pi \rfloor + 1}{n}, \tau\right) - L_n\left(\frac{\lceil n\pi \rfloor}{n}, \tau\right)\right].$$

We first prove that the continuous process $\{L_n(\cdot,\cdot)\}$ satisfies the Csensov tightness criterion, and then obtain the tightness of $\{\overline{S}_n(\cdot,\cdot)\}$ by proving the contiguity of L_n and $\overline{S}_n$. See Van der Vaart (1998, pp.86-91).

Substep 3a. We prove the tightness of $\{L_n(\cdot,\cdot)\}$. Noting that $L_n(0,0)=L_n(0,\tau)=0$, the first two conditions of the Csensov criterion are automatically satisfied. We now prove that $\{L_n(\cdot,\underline{\tau})\}$ is tight in $\mathcal{C}([0,1];\mathbb{R})$. Let $v_i(\tau)\equiv E[G(\varepsilon_{i\tau}/\lambda_n)]-G(\varepsilon_{i\tau}/\lambda_n)$, $v_i\equiv v_i(\underline{\tau})$, $\varsigma_{ij}(\tau)\equiv c_0 f_i(X_j)\,\overline{f}_n^{-1}(X_j)$, and $\varsigma_{ij}\equiv\varsigma_{ij}(\underline{\tau})$. Define

$$\overrightarrow{S}_n(\pi,\tau)=n^{-1/2}\sum_{i=1}^{\lceil n\pi\rfloor}v_i(\tau)-n^{-3/2}\sum_{i=1}^{\lceil n\pi\rfloor}\sum_{j=1}^{n}\varsigma_{ij}v_j(\tau).$$

Noting that $L_n(k/n,\tau)=\overrightarrow{S}_n(k/n,\tau)$ and $(n\pi-\lceil n\pi\rfloor)[L_n((\lceil n\pi\rfloor+1)/n,\tau)-L_n(\lceil n\pi\rfloor/n,\tau)]=o_P(1)$ uniformly in π (and τ), we can prove the tightness of $\{L_n(\cdot,\underline{\tau})\}$ in $\mathcal{C}([0,1];\mathbb{R})$ by checking the tightness of $\overrightarrow{S}_n(\cdot,\underline{\tau})$ in $\mathcal{D}([0,1];\mathbb{R})$ where $\mathcal{D}([0,1];\mathbb{R})$ is the space of real-valued processes indexed by $\pi\in[0,1]$ that are *cadlag* (right continuous with left limits). Let $0\le\pi_1<\pi<\pi_2\le 1$. By the Cauchy-Schwarz inequality,

$$E\left\{\left[\overrightarrow{S}_n(\pi,\underline{\tau})-\overrightarrow{S}_n(\pi_1,\underline{\tau})\right]^2\left[\overrightarrow{S}_n(\pi_2,\underline{\tau})-\overrightarrow{S}_n(\pi,\underline{\tau})\right]^2\right\}\le b_{n1}+b_{n2}+b_{n3}+b_{n4}$$

where

$$\begin{aligned}
b_{n1}&\equiv 4n^{-2}E\left\{\left[\sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}v_i\right]^2\left[\sum_{i=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}v_i\right]^2\right\},\\
b_{n2}&\equiv 4n^{-6}E\left\{\left[\sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}\sum_{j=1}^{n}\varsigma_{ij}v_j\right]^2\left[\sum_{i=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}\sum_{j=1}^{n}\varsigma_{ij}v_j\right]^2\right\},\\
b_{n3}&\equiv 4n^{-4}E\left\{\left[\sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}v_i\right]^2\left[\sum_{i=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}\sum_{j=1}^{n}\varsigma_{ij}v_j\right]^2\right\},\text{ and}\\
b_{n4}&\equiv 4n^{-4}E\left\{\left[\sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}\sum_{j=1}^{n}\varsigma_{ij}v_j\right]^2\left[\sum_{i=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}v_i\right]^2\right\}.
\end{aligned}$$

$$\begin{aligned}
b_{n1}&=4n^{-2}E\sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}\sum_{j=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}\sum_{k=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}\sum_{l=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}E(v_iv_jv_kv_l)\\
&=4n^{-2}\sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}\sum_{k=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}E\left(v_i^2v_k^2\right)+16n^{-2}\sum_{\lceil n\pi_1\rfloor+1\le i<j\le\lceil n\pi\rfloor}\sum_{\lceil n\pi\rfloor+1\le k<l\le\lceil n\pi_2\rfloor}E(v_iv_jv_kv_l)\\
&\quad+8n^{-2}\sum_{\lceil n\pi_1\rfloor+1\le i<j\le\lceil n\pi\rfloor}\sum_{k=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}E\left(v_iv_jv_k^2\right)+8n^{-2}\sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}\sum_{\lceil n\pi\rfloor+1\le k<l\le\lceil n\pi_2\rfloor}E\left(v_i^2v_kv_l\right)\\
&\equiv b_{n11}+b_{n12}+b_{n13}+b_{n14}.
\end{aligned}$$

Since $\max_{1\leq i\leq n}|v_i| \leq 1$, we have $b_{n11} \leq 4\{(\lceil n\pi\rfloor - \lceil n\pi_1\rfloor)/n\}\{(\lceil n\pi_2\rfloor - \lceil n\pi\rfloor)/n\}$. For b_{n12}, we consider two cases: (a) $l-k \leq j-i$, (b) $j-i < l-k$. In case (a), we apply the Davydov inequality to obtain

$$\begin{aligned} b_{n12a} &\equiv 16n^{-2} \sum_{\lceil n\pi_1\rfloor+1\leq i<j\leq\lceil n\pi\rfloor} \sum_{\substack{\lceil n\pi\rfloor+1\leq k<l\leq\lceil n\pi_2\rfloor \\ l-k\leq j-i}} E\left(v_i v_j v_k v_l\right) \\ &\leq 64n^{-2} \sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor} \sum_{j=i+1}^{\lceil n\pi\rfloor} \sum_{k=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor} (j-i)\,\alpha\,(j-i) \\ &\leq 64\frac{\lceil n\pi\rfloor - \lceil n\pi_1\rfloor}{n}\frac{\lceil n\pi_2\rfloor - \lceil n\pi\rfloor}{n}\sum_{s=1}^{\infty} s\alpha\,(s) \leq C(\pi-\pi_1)(\pi_2-\pi). \end{aligned}$$

Similar inequality holds true for case (b). Thus $b_{n12} \leq C(\pi-\pi_1)(\pi_2-\pi)$. Next,

$$\begin{aligned} b_{n13} &= 8n^{-2} \sum_{\lceil n\pi_1\rfloor+1\leq i<j\leq\lceil n\pi\rfloor} \sum_{k=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor} E\left(v_i v_j v_k^2\right) \\ &\leq 32n^{-2} \sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor} \sum_{k=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor} \sum_{s=1}^{\infty} \alpha\,(s) \leq C(\pi-\pi_1)(\pi_2-\pi). \end{aligned}$$

Similarly, $b_{n14} \leq C(\pi-\pi_1)(\pi_2-\pi)$. Consequently we have

$$b_{n1} \leq C(\pi-\pi_1)(\pi_2-\pi). \tag{B.3}$$

To find an upper bound for b_{n2}, we first apply the Cauchy-Schwarz inequality to obtain

$$\begin{aligned} b_{n2} &\leq \left\{4n^{-6}E\left[\sum_{i=\lceil n\pi_1\rfloor+1}^{\lceil n\pi\rfloor}\sum_{j=1}^{n}\varsigma_{ij}v_j\right]^4\right\}^{1/2}\left\{4n^{-6}E\left[\sum_{i=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}\sum_{j=1}^{n}\varsigma_{ij}v_j\right]^4\right\}^{1/2} \\ &\equiv \{b_{n21}\}^{1/2}\{b_{n22}\}^{1/2}. \end{aligned}$$

Let $\xi_j \equiv \sum_{i=\lceil n\pi\rfloor+1}^{\lceil n\pi_2\rfloor}\varsigma_{ij}$, where we suppress the dependence of ξ_j on n, π and π_2. Let $\|\xi\|_s \equiv \{E|\xi|^s\}^{1/s}$ for $s \geq 1$. Then by the Davydov inequality, the Hölder inequality, and Assumption A1, we can use the same trick as in the proof of b_{n12} to obtain

$$\begin{aligned} b_{n22} &= 4n^{-6}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\sum_{j_3=1}^{n}\sum_{j_4=1}^{n} E\left[\xi_{j_1}\xi_{j_2}\xi_{j_3}\xi_{j_4}\right] \\ &\leq 96n^{-6}\sum_{1\leq j_1\leq j_2\leq j_3\leq j_4\leq n}^{n}\left|E\left[\xi_{j_1}\xi_{j_2}\xi_{j_3}\xi_{j_4}\right]\right| \\ &\leq Cn^{-6}\sup_{n\geq1}\max_{1\leq j_1\leq j_2\leq j_3\leq j_4\leq n}\left\|\xi_{j_1}\right\|_{4+\eta}\left\|\xi_{j_2}\xi_{j_3}\xi_{j_4}\right\|_{(4+\eta)/3}\sum_{1\leq j_1\leq j_2\leq j_3\leq n}^{n}(j_2-j_1)\,\alpha\,(j_2-j_1)^{\eta/(4+\eta)} \\ &\leq Cn^{-4}\sup_{n\geq1}\max_{1\leq j\leq n}\left\|\xi_j\right\|_{4+\eta}^4\sum_{s=0}^{\infty}s\alpha\,(s)^{\eta/(4+\eta)} \leq Cn^{-4}\sup_{n\geq1}\max_{1\leq j\leq n}\left\|\xi_j\right\|_{4+\eta}^4. \end{aligned}$$

Then by the definition of ξ_j, the Minkowski inequality, and Assumptions A3(i)-(ii),

$$b_{n22} \leq C \left\{ \sup_{n\geq 1} \max_{1\leq j\leq n} n^{-1} \left\| \sum_{i=\lceil n\pi \rfloor+1}^{\lceil n\pi_2 \rfloor} \varsigma_{ij} \right\|_{4+\eta} \right\}^4 \leq C \left\{ \sup_{n\geq 1} \max_{1\leq j\leq n} n^{-1} \sum_{i=\lceil n\pi \rfloor+1}^{\lceil n\pi_2 \rfloor} \|\varsigma_{ij}\|_{4+\eta} \right\}^4 \leq C (\pi_2 - \pi)^4$$

where, again, the exact value of C varies across lines. Analogously, we can show that $b_{n21} \leq C(\pi - \pi_1)^4$. Then $b_{n2} \leq C(\pi_2 - \pi_1)^4$. Similarly, one can show that $b_{nl} \leq C(\pi_2 - \pi_1)^3$, $l = 3, 4$. It follows that $E\{[\overrightarrow{S}_n(\pi, \underline{\tau}) - \overrightarrow{S}_n(\pi_1, \underline{\tau})]^2 \; [\overrightarrow{S}_n(\pi_2, \underline{\tau}) - \overrightarrow{S}_n(\pi, \underline{\tau})]^2\} \leq C(\pi - \pi_1)(\pi_2 - \pi)$. The tightness of $\{L_n(\cdot, \underline{\tau})\}$ in $\mathcal{C}([0,1]; \mathbb{R})$ thus follows.

We now verify the last condition of the Csensov criterion for the process $\{L_n(\cdot,\cdot)\}$. Let $L_n(B) \equiv [L_n(\pi_2,\tau_2) - L_n(\pi_1,\tau_2)] - [L_n(\pi_2,\tau_1) - L_n(\pi_1,\tau_1)] \equiv \xi_{\pi_2\pi_1}(\tau_2) - \xi_{\pi_2\pi_1}(\tau_1)$. Our aim is to control $E|L_n(B)|^2$. Noting that

$$\begin{aligned}
\xi_{\pi_2\pi_1}(\tau_2) &= \left[n^{-1/2} \sum_{i=\lceil n\pi_1 \rfloor+1}^{\lceil n\pi_2 \rfloor} v_i(\tau_2) - n^{-3/2} \sum_{i=\lceil n\pi_1 \rfloor+1}^{\lceil n\pi_2 \rfloor} \sum_{j=1}^{n} \varsigma_{ij} v_j(\tau_2) \right] \\
&\quad + (n\pi_2 - \lceil n\pi_2 \rfloor) \left[n^{-1/2} v_{\lceil n\pi_2 \rfloor+1}(\tau_2) - n^{-3/2} \sum_{j=1}^{n} \varsigma_{\lceil n\pi_2 \rfloor+1,j} v_j(\tau_2) \right] \\
&\quad - (n\pi_1 - \lceil n\pi_1 \rfloor) \left[n^{-1/2} v_{\lceil n\pi_1 \rfloor+1}(\tau_2) - n^{-3/2} \sum_{j=1}^{n} \varsigma_{\lceil n\pi_1 \rfloor+1,j} v_j(\tau_2) \right] \\
&\equiv \vartheta_1(\pi_1,\pi_2,\tau_2) + \vartheta_2(\pi_1,\pi_2,\tau_2) - \vartheta_3(\pi_1,\pi_2,\tau_2), \text{ say,}
\end{aligned}$$

we have

$$E|L_n(B)|^2 \leq 3 \sum_{j=1}^{3} E[\vartheta_j(\pi_1,\pi_2,\tau_2) - \vartheta_j(\pi_1,\pi_2,\tau_1)]^2. \tag{B.4}$$

Noting that $\overline{\lim}_{n\to\infty} \|v_i(\tau_2) - v_i(\tau_1)\|_{2+\eta} \leq 2\overline{\lim}_{n\to\infty} \left\{ E|G(\varepsilon_{i\tau_2}/\lambda_n) - G(\varepsilon_{i\tau_1}/\lambda_n)|^{2+\eta} \right\}^{1/(2+\eta)} \leq C(\tau_2 - \tau_1)^{1/(2+\eta)}$ and $\overline{\lim}_{n\to\infty} \|\varsigma_{ij} v_j(\tau_2) - \varsigma_{ij} v_j(\tau_1)\|_{2+\eta} \leq c_f \overline{\lim}_{n\to\infty} \|v_j(\tau_2) - v_j(\tau_1)\|_{2+\eta} \leq C$ $|\tau_2 - \tau_1|^{1/(2+\eta)}$ where $c_f \equiv c_0 \overline{\lim}_{n\to\infty} \max_{1\leq i\leq n} \sup_{x\in\mathcal{X}} f_i(x) \overline{f}_n^{-1}(x) < \infty$, we have

$$\begin{aligned}
&\overline{\lim}_{n\to\infty} E[\vartheta_1(\pi_1,\pi_2,\tau_2) - \vartheta_1(\pi_1,\pi_2,\tau_1)]^2 \\
\leq\ & 2\overline{\lim}_{n\to\infty} \left\{ n^{-1} E\left(\sum_{i=\lceil n\pi_1 \rfloor+1}^{\lceil n\pi_2 \rfloor} [v_i(\tau_2) - v_i(\tau_1)] \right)^2 + n^{-3} E\left(\sum_{i=\lceil n\pi_1 \rfloor+1}^{\lceil n\pi_2 \rfloor} \sum_{j=1}^{n} \overline{\varsigma}_{ij}(\tau_1,\tau_2) \right)^2 \right\} \\
\leq\ & \overline{\lim}_{n\to\infty} 8 \frac{\lceil n\pi_2 \rfloor - \lceil n\pi_1 \rfloor}{n} \|v_i(\tau_2) - v_i(\tau_1)\|_{2+\eta}^2 \sum_{s=1}^{\infty} \alpha(s)^{\eta/(2+\eta)} \\
& + \overline{\lim}_{n\to\infty} 8 \frac{\lceil n\pi_2 \rfloor - \lceil n\pi_1 \rfloor}{n} \|\overline{\varsigma}_{ij}(\tau_1,\tau_2)\|_{2+\eta}^2 \sum_{s=1}^{\infty} s\alpha(s)^{\eta/(2+\eta)} \\
\leq\ & C(\pi_2 - \pi_1)|\tau_2 - \tau_1|^{2/(2+\eta)},
\end{aligned} \tag{B.5}$$

where $\overline{\varsigma}_{ij}(\tau_1,\tau_2) \equiv \varsigma_{ij}[v_j(\tau_2) - v_j(\tau_1)]$. Noting that $\sup_{\pi\in[0,1]} |n\pi - \lceil n\pi \rfloor| = 1$, it is easy to show that

$$E[\vartheta_l(\pi_1,\pi_2,\tau_2) - \vartheta_l(\pi_1,\pi_2,\tau_1)]^2 \leq C\frac{\tau_2 - \tau_1}{n} \text{ for } l = 2, 3. \tag{B.6}$$

Combining (B.4)-(B.6) yields $\overline{\lim}_{n\to\infty} E\left|L_n(B)\right|^2 \leq C(\pi_2-\pi_1)\left|\tau_2-\tau_1\right|^{2/(2+\eta)}$ for large enough C. This completes the proof of the tightness of $\{L_n(\cdot,\cdot)\}$ in $\mathcal{C}([0,1]\times\mathcal{T};\mathbb{R})$.

Substep 3b. We prove that $\{L_n(\cdot,\cdot)\}$ and $\{\overline{S}_n(\cdot,\cdot)\}$ are contiguous by proving that for any $\epsilon>0$, $P\left(\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}\left|L_n(\pi,\tau)-\overline{S}_n(\pi,\tau)\right|>2\epsilon\right)\to 0$ as $n\to\infty$. Note that

$$\begin{aligned}
&P\left(\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}\left|L_n(\pi,\tau)-\overline{S}_n(\pi,\tau)\right|>2\epsilon\right)\\
&\leq P\left(\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}\left|L_n(\pi,\tau)-\overrightarrow{S}_n(\pi,\tau)\right|>\epsilon\right)+P\left(\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}\left|\overrightarrow{S}_n(\pi,\tau)-\overline{S}_n(\pi,\tau)\right|>\epsilon\right). \quad \text{(B.7)}
\end{aligned}$$

Following the proof of Proposition A.3, we can readily show that the second term in (B.7) tends to 0 as $n\to\infty$. Noting that

$$\begin{aligned}
\left|L_n(\pi,\tau)-\overrightarrow{S}_n(\pi,\tau)\right| &= \left|n\pi-\lceil n\pi\rfloor\right| L_n\left(\left[\frac{\lceil n\pi\rfloor}{n},\frac{\lceil n\pi\rfloor+1}{n}\right]\times[\underline{\tau},\tau]\right)\\
&\leq L_n\left(\left[\frac{\lceil n\pi\rfloor}{n},\frac{\lceil n\pi\rfloor+1}{n}\right]\times[\underline{\tau},\tau]\right),
\end{aligned}$$

it follows from the Chebyshev inequality that

$$P\left(\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}\left|L_n(\pi,\tau)-\overrightarrow{S}_n(\pi,\tau)\right|\geq\epsilon\right)\leq\epsilon^{-2}E\left(\max_{1\leq k\leq n}\sup_{\tau\in\mathcal{T}}\left|L_n\left(\left[\frac{k}{n},\frac{k+1}{n}\right]\times[\underline{\tau},\tau]\right)\right|\right)^2. \quad \text{(B.8)}$$

Applying the fourth condition of the Csensov criterion to the process $\{L_n(\pi,\tau)\}$ with $B_k=[k/n,(k+1)/n]\times[\underline{\tau},\tau]$ yields

$$E\left|L_n(B_k)\right|^2\leq\frac{C(\tau-\underline{\tau})^{\eta/(2+\eta)}}{n}\text{ for each }k=1,2,\cdots,n\text{ and large }n. \quad \text{(B.9)}$$

By the continuity of the process $\{L_n(\pi,\tau)\}$ in $\tau\in\mathcal{T}$, there exists τ_k such that

$$\sup_{\tau\in\mathcal{T}}\left|L_n(B_k)\right|=L_n\left(\left[\frac{k}{n},\frac{k+1}{n}\right]\times[\underline{\tau},\tau_k]\right),$$

Then there exists k^* such that

$$\begin{aligned}
\max_{1\leq k\leq n}\sup_{\tau\in\mathcal{T}}\left|L_n\left(\left[\frac{k}{n},\frac{k+1}{n}\right]\times[\underline{\tau},\tau_k]\right)\right| &= \max_{1\leq k\leq n}\left|L_n\left(\left[\frac{k}{n},\frac{k+1}{n}\right]\times[\underline{\tau},\tau_k]\right)\right|\\
&= \left|L_n\left(\left[\frac{k^*}{n},\frac{k^*+1}{n}\right]\times[\underline{\tau},\tau_{k^*}]\right)\right|.
\end{aligned}$$

This, together with (B.9), implies that

$$E\left(\max_{1\leq k\leq n}\sup_{\tau\in\mathcal{T}}\left|L_n\left(\left[\frac{k}{n},\frac{k+1}{n}\right]\times[\underline{\tau},\tau_k]\right)\right|\right)^2\leq\frac{C(\overline{\tau}-\underline{\tau})^{\eta/(2+\eta)}}{n}. \quad \text{(B.10)}$$

It follows from (B.8) and (B.10) that $P\left(\sup_{\pi\in[0,1]}\sup_{\tau\in\mathcal{T}}\left|L_n(\pi,\tau)-\overrightarrow{S}_n(\pi,\tau)\right|\geq\epsilon\right)\to 0$. That is, $\{L_n(\cdot,\cdot)\}$ and $\{\overline{S}_n(\cdot,\cdot)\}$ are contiguous. ■

Proof of Corollary 3.2

Under Assumption A8(i), $\overline{f}_{\lceil n\pi \rfloor}(X_t) - \pi \overline{f}_n(X_t) = o_{a.s.}(1)$ uniformly in π. This, together with (B.1) and (B.2), implies that under H_{1n},

$$S_n(\pi,\tau) = \frac{1}{n^{1/2}} \sum_{i=1}^{\lceil n\pi \rfloor} \psi_\tau(\varepsilon_{i\tau}) - \frac{\pi c_0}{n^{1/2}} \sum_{i=1}^{n} \psi_\tau(\varepsilon_{i\tau}) + \Delta(\pi,\tau) + o_P(1), \tag{B.11}$$

where $o_P(1)$ holds uniformly in $(\pi,\tau) \in [0,1] \times \mathcal{T}$. The rest of the proof then follows directly from the calculation of the covariance kernel, Assumption A8(ii), and Theorem 3.1. ■

Proof of Theorem 3.3

By (B.11), under H_{1n}, we have uniformly in $(\pi,\tau) \in [0,1] \times \mathcal{T}$,

$$\begin{aligned} S_n^c(\pi,\tau) &= \left\{ \frac{1}{n^{1/2}} \sum_{i=1}^{\lceil n\pi \rfloor} \psi_\tau(\varepsilon_{i\tau}) - \frac{\pi}{n^{1/2}} \sum_{i=1}^{n} \psi_\tau(\varepsilon_{i\tau}) \right\} \\ &\quad + \left\{ \Delta^0(\pi,\tau) - \pi\Delta^0(1,\tau) \right\} + o_P(1), \end{aligned}$$

where $\Delta^0(\pi,\tau) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{\lceil n\pi \rfloor} E[f_0(m_{0t\tau}|X_t)\delta(\tau, X_t, t/n)] = \int_0^\pi \int f_0(m_0(\tau,x)|x) f(x) \delta(\tau,x,s)\, dxds$. The result then follows from Theorem 3.1 and the direct calculation of the limiting covariance kernel of $n^{-1/2}\sum_{i=1}^{\lceil n\pi \rfloor} \psi_\tau(\varepsilon_{i\tau}) - \pi n^{-1/2} \sum_{i=1}^{n} \psi_\tau(\varepsilon_{i\tau})$. ■

Proof of Corollary 3.4

This follows from Theorem 3.3 and the fact that $\Gamma^0(\tau_1,\tau_2) = \tau_1 \wedge \tau_2 - \tau_1\tau_2$ under the m.d.s. condition.■

Proof of Theorem 3.5

It suffices to show that $S_n^*(\cdot,\cdot) \overset{p}{\Rightarrow} S_\infty^0(\cdot,\cdot)$. Let P^* denote the probability conditional on the original sample $\mathcal{W} \equiv \{(Y_t, X_t)\}_{t=1}^n$. Let E^* denote the expectation with respect to P^*. To proceed, rewrite $S_n^*(\pi,\tau) = \sum_{i=1}^n s_{ni}(z_i;\pi,\tau)$, where

$$s_{ni}(z;\pi,\tau) = z\, \mathbf{1}(i \le \lceil n\pi \rfloor - l + 1)\, n^{-1/2} \sum_{j=i}^{i+l-1} [\tau - G(-\widehat{u}_{j\tau}/\lambda_n)]. \tag{B.12}$$

Define the envelope function of s_{ni} as

$$\overline{s}_{ni}(z_i) = |z_i|\, n^{-1/2} \sup_{\tau\in\mathcal{T}} \left| \sum_{j=i}^{i+l-1} [\tau - G(-\widehat{u}_{j\tau}/\lambda_n)] \right|. \tag{B.13}$$

Conditional on $\mathcal{W}$, the triangular array $\{s_{ni}(z_i;\pi,\tau)\}$ are independent within rows so that we can apply Theorem 10.6 of Pollard (1990) to show the weak convergence of $S_n^*(\cdot,\cdot)$ to $S_\infty(\cdot,\cdot)$. Define the pseudo-metric

$$\rho_n(\pi,\pi';\tau,\tau') = \left\{ \sum_{i=1}^n E^*\left[(s_{ni}(z_i;\pi',\tau') - s_{ni}(z_i;\pi,\tau))^2 \right] \right\}^{1/2}. \tag{B.14}$$

By Theorem 10.6 of Pollard (1990), it suffices to verify the following five conditions:

(i) $\{s_{ni}\}$ are manageable in the sense of Definition 7.9 of Pollard (1990);

(ii) $E^*\left[S_n^*(\pi,\tau)\, S_n^*(\pi',\tau')\right] \xrightarrow{P} (\pi\wedge\pi')\Gamma^0(\tau,\tau')$ for every (π,τ), (π',τ') in $[0,1]\times\mathcal{T}$;

(iii) $\overline{\lim}_{n\to\infty}\sum_{i=1}^n E^*\left(\overline{s}_{ni}^2(z_i)\right)$ is stochastically bounded;

(iv) $\sum_{i=1}^n E^*\left(\overline{s}_{ni}^2(z_i)\ \mathbf{1}\left(\overline{s}_{ni}(z_i)>\epsilon\right)\right) \xrightarrow{P} 0$ for each $\epsilon>0$;

(v) $\rho(\pi,\pi';\tau,\tau') \equiv$plim$_{n\to\infty}\rho_n(\pi,\pi';\tau,\tau')$ is well defined and, for all deterministic sequences $\{\pi_n',\tau_n'\}$ and $\{\pi_n,\tau_n\}$, if $\rho(\pi_n,\pi_n';\tau_n,\tau_n')\to 0$ then $\rho_n(\pi_n,\pi_n';\tau_n,\tau_n')\xrightarrow{P} 0$.

Step 1. We verify condition (i). In order for the triangular array of process $\{s_{ni}(z_i;\pi,\tau)\}$ to be manageable with respect to the envelope $\overline{s}_{ni}(z_i)$, we need to find a deterministic function $\lambda(\epsilon_0)$ that bounds the the covering number of $\boldsymbol{\alpha}\odot\mathbf{S}_n \equiv \{\alpha_i s_{ni}(z_i;\pi,\tau) : \pi\in[0,1],\ \tau\in\mathcal{T},\ \alpha_i$ are nonnegative finite constants for all $i=1,\cdots n\}$ with $\sqrt{\log\lambda(\epsilon_0)}$ integrable. Here, the covering number refers to the smallest number of closed balls with radius $(\epsilon_0/2)\sqrt{\sum_{i=1}^n\alpha_i^2\overline{s}_{ni}(z_i)^2}$ whose unions cover $\boldsymbol{\alpha}\odot\mathbf{S}_n$. It follows that within each closed ball

$$\sum_{i=1}^n \alpha_i^2 E^*\left[s_{ni}(z_i;\pi,\tau)-s_{ni}(z_i;\pi',\tau')\right]^2 \le \frac{\epsilon_0^2}{4}\sum_{i=1}^n\alpha_i^2 E^*\left[\overline{s}_{ni}(z_i)^2\right]\ \ \forall\epsilon_0\in(0,1]. \tag{B.15}$$

First we study the term on the left hand side (l.h.s.) of (B.15). Without loss of generality (W.l.o.g.), assume that $\pi<\pi'$. Then

$$\sum_{i=1}^n \alpha_i^2 E^*\left[s_{ni}(z_i;\pi,\tau)-s_{ni}(z_i;\pi',\tau')\right]^2 = \frac{1}{n}\sum_{i=1}^{\lceil n\pi\rfloor-l+1}\alpha_i^2\mathbb{Z}_{1i} + \frac{1}{n}\sum_{i=\lceil n\pi\rfloor-l+2}^{\lceil n\pi'\rfloor-l+1}\alpha_i^2\mathbb{Z}_{2i},$$

where

$$\mathbb{Z}_{1i}\equiv\frac{1}{l}\left\{\sum_{j=i}^{i+l-1}\left[\tau-G(-\widehat{u}_{j\tau}/\lambda_n)-\tau'+G(-\widehat{u}_{j\tau'}/\lambda_n)\right]\right\}^2,\ \text{and}\ \mathbb{Z}_{2i}\equiv\frac{1}{l}\left\{\sum_{j=i}^{i+l-1}\left[\tau'-G(-\widehat{u}_{j\tau'}/\lambda_n)\right]\right\}^2.$$

By Propositions A.2 and A.6,

$$\begin{aligned}\frac{1}{\sqrt{l}}\sum_{j=i}^{i+l-1}\left[\tau-G(-\widehat{u}_{j\tau}/\lambda_n)\right] &= \frac{1}{\sqrt{l}}\sum_{j=i}^{i+l-1}\psi_\tau(\varepsilon_{j\tau}) + l^{-1/2}\sum_{j=i}^{i+l-1}\left[\mathbf{1}(\varepsilon_{j\tau}<0)-G(-\widehat{u}_{j\tau}/\lambda_n)\right]\\ &= \frac{1}{\sqrt{l}}\sum_{j=i}^{i+l-1}\psi_\tau(\varepsilon_{j\tau})+o_P(1)\ \text{ uniformly in }(\pi,\tau).\end{aligned} \tag{B.16}$$

It follows that

$$\begin{aligned}\mathbb{Z}_{1i} &= \left\{\frac{1}{\sqrt{l}}\sum_{j=i}^{i+l-1}\left[\psi_\tau(\varepsilon_{j\tau})-\psi_{\tau'}(\varepsilon_{j\tau'})\right]\right\}^2+o_P(1)\xrightarrow{P}\overline{\Gamma}(\tau,\tau'),\ \text{and}\\ \mathbb{Z}_{2i} &= \left\{\frac{1}{\sqrt{l}}\sum_{j=i}^{i+l-1}\psi_{\tau'}(\varepsilon_{j\tau'})\right\}^2+o_P(1)\xrightarrow{P}\Gamma^0(\tau',\tau'),\end{aligned}$$

where $\overline{\Gamma}(\tau,\tau')=\Gamma^0(\tau,\tau)-2\Gamma^0(\tau,\tau')+\Gamma^0(\tau',\tau')$. It follows that when $\pi<\pi'$, we have

$$\sum_{i=1}^n\alpha_i^2E^*\left[s_{ni}(z_i;\pi,\tau)-s_{ni}(z_i;\pi',\tau')\right]^2\xrightarrow{P}\sum_{i=1}^\infty\alpha_i^2\left[\pi\Gamma^0(\tau,\tau)-2\pi\Gamma^0(\tau,\tau')+\pi'\Gamma^0(\tau',\tau')\right].$$

And for generic $\pi, \pi' \in [0,1]$, we have

$$\sum_{i=1}^{n} \alpha_i^2 E^* \left[s_{ni}(z_i; \pi, \tau) - s_{ni}(z_i; \pi', \tau')\right]^2 \xrightarrow{P} \sum_{i=1}^{\infty} \alpha_i^2 \rho^2(\pi, \pi', \tau, \tau'), \tag{B.17}$$

where $\rho^2(\pi, \pi', \tau, \tau') \equiv \pi\Gamma^0(\tau, \tau) - 2(\pi \wedge \pi')\Gamma^0(\tau, \tau') + \pi'\Gamma^0(\tau', \tau')$.

Next, we study the term on the right hand side (r.h.s.) of (B.15). By (B.16)

$$\sum_{i=1}^{n} \alpha_i^2 E^* \left(\overline{s}_{ni}^2(z_i)\right) \leq \frac{1}{n} \sum_{i=1}^{n} \alpha_i^2 \sup_{\pi \in [0,1]} \sup_{\tau \in \mathcal{T}} \left\{ l^{-1/2} \sum_{j=i}^{i+l-1} \psi_\tau(\varepsilon_{j\tau}) \right\}^2 + o_P(1) = O_P(1), \tag{B.18}$$

where the last equality follows because $\{l^{-1/2} \sum_{j=i}^{i+l-1} \psi_\tau(\varepsilon_{j\tau})\}$ is an empirical process indexed by τ that satisfies the Donsker theorem. This, together with (B.15) and (B.17), implies that for any small $\epsilon_1 > 0$, there exists a large constant $M_1 \equiv M_1(\epsilon_1)$ such that the following holds

$$\sum_{i=1}^{\infty} \alpha_i^2 \rho^2(\pi, \pi'; \tau, \tau') \leq \frac{\epsilon_0^2}{4} M_1 \text{ for sufficiently large } n \tag{B.19}$$

on a set with probability $1 - \epsilon_1$.

Now, let $\pi_j = j\sigma_1$ for $j = 0, 1, 2, \cdots$. We partition the compact set $\mathcal{T}$ by finite points $\underline{\tau} = \tau_0 < \tau_1 < \cdots < \tau_{N2-1} < \tau_{N2} = \overline{\tau}$ such that $|\tau_j - \tau_{j-1}| = \sigma_2$. W.l.o.g., set $\sigma_2 = \sigma_1 = \sigma \in (0,1)$. Let $(\pi, \tau) \in [\pi_{j-1}, \pi_j] \times [\tau_{k-1}, \tau_k]$ $(j, k = 1, 2, \cdots)$. Note $\rho^2(\pi_j, \pi, \tau_k, \tau) = \pi\overline{\Gamma}(\tau_k, \tau) + (\pi_j - \pi)\Gamma^0(\tau_k, \tau_k)$. By the fact that $\left\|\psi_\tau(\varepsilon_{i\tau}) - \psi_{\tau_k}(\varepsilon_{i\tau_k})\right\|_{2+\eta}^{2+\eta} \leq C\sigma$ and Davydov inequality,

$$\overline{\Gamma}(\tau_k, \tau) = \lim_{n \to \infty} \frac{1}{n} E\left\{\sum_{i=1}^{n} \left[\psi_\tau(\varepsilon_{i\tau}) - \psi_{\tau_k}(\varepsilon_{i\tau_k})\right]\right\}^2 \leq C\sigma^{2/(2+\eta)} \sum_{s=1}^{\infty} \alpha(s)^{1/(2+\eta)} \leq C\sigma^{2/(2+\eta)},$$

where the exact value of C varies across lines. It follows that

$$\rho^2(\pi_j, \pi; \tau_k, \tau) \leq C\sigma^{2/(2+\eta)} + \sigma \max_k \Gamma^0(\tau_k, \tau_k) \leq C_1\sigma^{2/(2+\eta)}$$

for large enough C_1. Consequently, if we choose $\sigma = \epsilon_0^{2+\eta}$, then $\sum_{i=1}^{\infty} \alpha_i^2 \rho^2(\pi_j, \pi; \tau_k, \tau) \leq C_1\epsilon_0^2 \sum_{i=1}^{\infty} \alpha_i^2$ so that (B.19) can be satisfied for sufficiently large n and M_1. It follows that the capacity bound is $O\left(\sigma^{-2}\right) = O(\epsilon_0^{-2(2+\eta)})$ and the integrability condition is satisfied.

Step 2. We verify condition (ii). By (B.16),

$$\begin{aligned}
& E^*\left[S_n^*(\pi, \tau) S_n^*(\pi', \tau')\right] \\
= \; & \frac{1}{n} \sum_{i=1}^{(\lceil n\pi \rfloor \wedge \lceil n\pi' \rfloor) - l + 1} \frac{1}{l} \sum_{j_1=i}^{i+l-1} \sum_{j_2=i}^{i+l-1} \left[\tau - G(-\widehat{u}_{j_1\tau}/\lambda_n)\right]\left[\tau' - G(-\widehat{u}_{j_2\tau'}/\lambda_n)\right] \\
= \; & \frac{1}{n} \sum_{i=1}^{(\lceil n\pi \rfloor \wedge \lceil n\pi' \rfloor) - l + 1} \frac{1}{l} \sum_{j_1=i}^{i+l-1} \sum_{j_2=i}^{i+l-1} \psi_\tau(\varepsilon_{j\tau}) \psi_{\tau'}(\varepsilon_{j\tau'}) + o_P(1) \\
\equiv \; & \overline{S}_n^*(\pi, \pi'; \tau, \tau') + o_P(1).
\end{aligned}$$

Let $\overline{S}_n^* \equiv \overline{S}_n^*(\pi,\pi';\tau,\tau')$. Then $E(\overline{S}_n^*) \to (\pi\wedge\pi')\Gamma^0(\tau,\tau')$. To show $\mathrm{Var}(\overline{S}_n^*) = o(1)$, let $\xi_{ni}^* \equiv \xi_{ni}^*(\tau,\tau') = \frac{1}{l^2}\sum_{j_1=i}^{i+l-1}\sum_{j_2=i}^{i+l-1}\psi_\tau(\varepsilon_{j_1\tau})\psi_{\tau'}(\varepsilon_{j_2\tau'})$, and let $\xi_{ni}(\tau) \equiv \frac{1}{l}\sum_{j=i}^{i+l-1}\psi_\tau(\varepsilon_{j\tau})$. Then by the Cauchy inequality,

$$\|\xi_{ni}^*\|_8 = \|\xi_{ni}(\tau)\xi_{ni}(\tau')\|_8 \le \|\xi_{ni}(\tau)\|_{16}\|\xi_{ni}(\tau')\|_{16}.$$

By Lemma 3.1 of Andrews and Pollard (1994) with Q =16, $\|\xi_{ni}(\tau)\|_{16}^{16} = E[\frac{1}{l}\sum_{j=i}^{i+l-1}\psi_\tau(\varepsilon_{j\tau})]^{16} = O(l^{-8})$. Consequently, $E(\xi_{ni}^{*8}) = O(l^{-8})$. Let $\kappa_{4n} = \sup_{i\le n}\sup_{\pi,\pi';\tau,\tau'} E[\xi_{ni}^{*8}] = O(l^{-8})$ and $\kappa_{2n} = \sup_{i\le n}\sup_{\pi,\pi';\tau,\tau'} E[\xi_{ni}^{*4}] = O(l^{-4})$. By Lemma A.1(b) of Inoue (2001) with $\delta = 2$,

$$E\left(\frac{l}{n}\sum_{i=1}^{n-l+1}\xi_{ni}^*\right)^4 = O\left(l^4 n^{-4} l^2\left(n^2\kappa_{4n}^{1/2} + n\kappa_{2n}\right)\right) = O\left(n^{-2}l^2\right) = o(1).$$

Hence $\overline{S}_n^* = (\pi\wedge\pi')\Gamma^0(\tau,\tau') + o_P(1)$ by the Chebyshev inequality, and $E^*[S_n^*(\pi,\tau)S_n^*(\pi',\tau')] \xrightarrow{P} (\pi\wedge\pi')\Gamma^0(\tau,\tau')$.

Step 3. We verify condition (iii). By choosing $\alpha_i = 1\ \forall i$ in (B.18), we have shown $\overline{\lim}_{n\to\infty}\sum_{i=1}^n E^*\left(\overline{s}_{ni}^2(z_i)\right)$ is stochastically bounded.

Step 4. We verify condition (iv). By the conditional Chebyshev inequality and (B.16),

$$\begin{aligned} P^*(\overline{s}_{ni}(z_i) > \epsilon) &\le \frac{l}{n\epsilon^2}\left\{\sup_{\tau\in\mathcal{T}}\frac{1}{l}\left|\sum_{j=i}^{i+l-1}(\tau - G(-\widehat{u}_{j\tau}/\lambda_n))\right|\right\}^2 \qquad \text{(B.20)} \\ &\le \frac{l}{n\epsilon^2}\left\{\sup_{\tau\in\mathcal{T}}\left|\frac{1}{\sqrt{l}}\sum_{j=i}^{i+l-1}\psi_\tau(\varepsilon_{i\tau})\right|^2 + o_P(1)\right\} = O_P\left(\frac{l}{n}\right). \end{aligned}$$

By the Cauchy-Schwarz inequality, (B.16) and (B.20),

$$\begin{aligned} &\sum_{i=1}^n E^*\left(\overline{s}_{ni}^2(z_i)\mathbf{1}(\overline{s}_{ni}(z_i) > \epsilon)\right) \\ =\ & \frac{1}{n}\sum_{i=1}^n E^*\left\{z_i^2\sup_{\tau\in\mathcal{T}}\left|\sum_{j=i}^{i+l-1}[\tau - G(-\widehat{u}_{j\tau}/\lambda_n)]\right|^2\mathbf{1}(\overline{s}_{ni}(z_i) > \epsilon)\right\} \\ \le\ & \frac{1}{n}\sum_{i=1}^n\left\{\sup_{\tau\in\mathcal{T}}\left|\frac{1}{\sqrt{l}}\sum_{j=i}^{i+l-1}\psi_\tau(\varepsilon_{i\tau}) + o_P(1)\right|^4 P^*(\overline{s}_{ni}(z_i) > \epsilon)\right\}^{1/2} \\ =\ & O_P\left(\sqrt{l/n}\right) = o_P(1). \end{aligned}$$

The result follows.

Step 5. We verify condition (v). From the verification of condition (i), we know that $\rho^2(\pi,\pi';\tau,\tau') = \mathrm{plim}_{n\to\infty}\rho_n^2(\pi,\pi';\tau,\tau')$ is well defined. If $\rho(\pi_n,\pi_n';\tau_n,\tau_n') \to 0$, then $\rho_n(\pi_n,\pi_n';\tau_n,\tau_n') \le |\rho_n(\pi_n,\pi_n';\tau_n,\tau_n') - \rho(\pi_n,\pi_n';\tau_n,\tau_n')| + \rho(\pi_n,\pi_n';\tau_n,\tau_n') \xrightarrow{P} 0$. ■